# A systems genomics approach to identify risk loci and pathways to Candida infection

Matzaraki, Vasiliki

A systems genomics approach

**to identify risk loci and pathways**

# to *Candida* infection

*Vasiliki Matzaraki*

**Vasiliki Matzaraki**

# A system genomics approach to identify risk loci and pathways to *Candida* infection

Thesis, University of Groningen, with summary in English and Dutch
The research described in this thesis was conducted at the Department of Genetics, University Medical Center Groningen, University of Groningen, The Netherlands

**Printing of this thesis was financially supported by the University of Groningen.**

Cover Design and Thesis Layout:


Printing: Gilderprint

ISBN: 978-94-6323-509-9



university of groningen

umcg

university of groningen

# A systems genomics approach to identify risk loci and pathways to *Candida* infection

**PhD thesis**

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Wednesday 6 March 2019 at 11.00 hours

by

**Vasiliki Matzaraki**

born on 14 April 1984
in Xanthi, Greece

**Supervisor**
Prof. T.N. Wijmenga

**Co-supervisor**
Prof. V.K. Magadi Gopalaiah

**Assessment Committee**
Prof. G. Molema
Prof. J.N.G. Oude Elberink
Prof. R. van Crevel

**Paranymphs:**
Arnau Vich Vila
Raúl Aguirre-Gamboa

# Table of Contents

# CHAPTER 1

## INTRODUCTION AND OUTLINE OF THE THESIS

Matzaraki V.

# CHAPTER 1

## INTRODUCTION AND OUTLINE OF THE THESIS

### • Human Genetic Susceptibility to Infectious Diseases

Infectious diseases are a major cause of human morbidity and mortality worldwide and, despite great advances in human medicine and genomics, they remain so, particularly in low income countries and in young children[1,2]. A main feature of human infectious diseases is that only a proportion of the individuals exposed to an infectious agent will actually develop disease, suggesting that genetic and health factors determine susceptibility to infectious diseases. Heritable factors were long considered to play a determining role in infectious studies and since 1923 the role of heredity has been studied by Webster and others using mice[3]. Indeed, single-gene (Mendelian) disorders leading to primary immuno¬deficiencies, although often rare, provided the first information on the role of genetics in determining susceptibility to infectious diseases. The best-known, relatively common, single-gene disorder that influences susceptibility to infection is the sickle cell heterozygosity against falciparum malaria[4]. Further clues came from early twin studies, which provided significant evidence on the role of human host genetic factors in diseases such as leprosy, poliomyelitis, hepatitis B, tuberculosis and *Helicopter* pylori infection[5–11]. In addition, follow-up studies of adopted children in the late 1980s showed they had a markedly increased risk to death from infection if one of the biological parents had died prematurely from an infection rather than other causes such as cancer or cardiovascular diseases[12]. Thus, an important role for host genetics in determining susceptibility to infectious diseases has been well established.

Moreover, genome-wide linkage studies provided further evidence of the role of genetics in infectious diseases, including *Schistosoma mansoni*, *H. pylori* and leprosy[13–17]. However, the main disadvantage of linkage studies is that it is difficult to recruit sufficient numbers of sibling pairs affected by the same infection. Thus, these studies are not sensitive enough to detect genes with a relatively small effect size that are characteristic for infectious diseases.

**Linkage studies:** Studies aimed at establishing linkage between genes. Linkage is the tendency of genes and genetic markers that are close together on a chromosome to be inherited together. These studies usually involve large families where the disease affects individuals in several generations.

**Sibling pairs:** Sibling pairs affected with the disease of interest are often the design of choice in linkage analysis studies with the aim to identifying genes that increase susceptibility to complex diseases. Such a design has the advantage of making few assumptions about the mode of inheritance of the disease.

## • Studying the Genetics of Infectious Diseases: What's New

Recent advances, such as a map of human genetic variation compiled by the international HapMap project[18] and the 1000 Genomes project[19], the application of genome-wide associations studies (GWAS) together with imputation tools (see *chapter 5*) and the development of new sequencing platforms, have all contributed to a better understanding of genetics in various human complex diseases. Such attempts have also focused on infectious diseases, although GWAS to identify genetic risk factors for these have not been as successful as in other complex diseases, such as immune-mediated diseases (Fig. 1A). As shown in Figure 1, there is a lack of genetic information for infectious diseases despite the fact that genetic variants associated with infectious diseases increase risk by 2-3 fold, an order of magnitude higher than observed for other complex diseases (Fig. 1B).

**Genome-wide association study (GWAS):** It is an approach that involves the genotyping of a genome-wide set of genetic variants in many individuals to find genetic variations associated with a particular disease. Such an approach is particularly useful in finding associations between single-nucleotide polymorphisms (SNPs) and common, complex diseases.

One of the major limitations in studying the genetics of infectious diseases is the lack of power due to small patient cohorts. The collection of such a cohort is complicated by the possible presence of co-morbidities in infected patients. Moreover, the use of appropriate controls is challenging as it is often unclear whether individuals have been exposed to the infectious agent due to the presence of asymptomatic infections. In addition to sufficient sample sizes and appropriate controls, another limitation of GWAS alone is that we cannot functionally interpret the identified associations and therefore need functional validation of novel findings from *in vitro* and/or *in vivo* experimental models.

Yet another limitation of studying infectious diseases stems from the emergence of an infectious disease, which is in principle the result of interaction by three main factors: pathogen, environment and human host genetics. Therefore, to better understand how individuals become infected, we need to capture this interaction but classical GWAS studies can detect only the genetic component. We require novel approaches to address the challenges of studying infectious diseases and highly integrated research that will offer a better understanding of the complex interplay between host, invading pathogen and environment. To this end, the development of high-throughput technologies, which allow us to collect multi-omics data (Table 1), together with advanced computational methods, has opened new opportunities to successfully implement a systems genomics approach. Instead of studying individual molecular components that contribute to a disease, we can

now generate and integrate high-throughput multi-omics data (genomics, transcriptomics, proteomics) in the context of the disease, which capture the dynamic interactions between pathogen-host-environment. In the next section, we will discuss the use of a systems genomics approach in the field of infectious diseases.

**Table 1.** Selected high-throughput methods for studying host-pathogen interactions.

| Method | Purpose | Reference |
|---|---|---|
| RNAseq | Transcript analysis | Nagalakshmi U. et al. 2008 |
| GRO-Seq | Transcription | Core L.J. et al. 2008 |
| PRO-seq | Genome-wide map of transcriptionally-engaged Pol II | Hojoong K. et al. 2013 |
| Nascent-Seq | Transcription | Khodor Y.L. et al. 2011 |
| ChIA-PET | Chromatin conformation | Fullwood M.J. et al. 2009 |
| Hi-C | Chromatin conformation | Lieberman A.E. et al. 2009; Rao et al. 2014 |
| 5-C-Seq | Chromatin conformation | Dotsie J. et al. 2006 |
| DNAse-Seq | Open chromatin | Crawford G.E. et al. 2006 |
| ATAC-Seq | Open chromatin | Buenrostro J.D. et al. 2013 |
| Chip-seq | Mapping DNA regulatory elements | Johnson D.S. et al. 2007 |
| BS-Seq | Genome methylation | Cokus S.J. et al. 2008 |
| RRBS-Seq | Genome methylation | Meissner A. et al. 2008 |
| ITS1-Seq | Fungi detection | Borman A.M. et al, 2008 |
| nanoLC-MS/MS | Host and fungal quantitative proteome analysis without isolation | Kitahara N. et al. 2015 |

Abbreviations: RNAseq RNA sequencing, GRO-Seq Global run-on sequencing, PRO-seq Precision nuclear Run-On and sequencing assay, ChIA-PET Chromatin interaction analysis by paired-end tag sequencing, ATAC-Seq Assay for transposase-accessible chromatin using sequencing, Chip-seq Chromatin immunoprecipitation sequencing, BS-Seq Bisulphite sequencing, RRBS-Seq Reduced representation bisulphite sequencing, ITS1-Seq Internal transcribed spacer region 1 sequencing, nanoLC-MS/MS Nanoscale liquid chromatoraphy coupled to tandem mass spectrometry

## • Systems Genomics Approach to Studying Host-Pathogen Interactions

Cancer research was one of the first fields to embrace the use of a systems genomics approach[20]. In response to the need for new approaches, the National Institute of Allergy and Infectious Diseases (NIAID, https://www.niaid.nih.gov/) has sponsored the Systems Biology Programme for Infectious Disease Research[21]. This programme consists of four centres, each of which focuses on studying interaction between the human host and a different bacterial or viral pathogen, including *Mycobacterium tuberculosis* bacillus, H5N1 avian influenza virus, and severe acute respiratory syndrome-associated coronavirus (SARS-CoV), influenza virus strains of various pathogenicities, *Salmonella enterica* serovar Typhimurium and *Yersinia* species.

One of the main advantages of a systems genomics approach is that multiple phenotypes, such as gene and protein expression data, can be collected for a reference population and subsequently associated with genetic variation. Using this

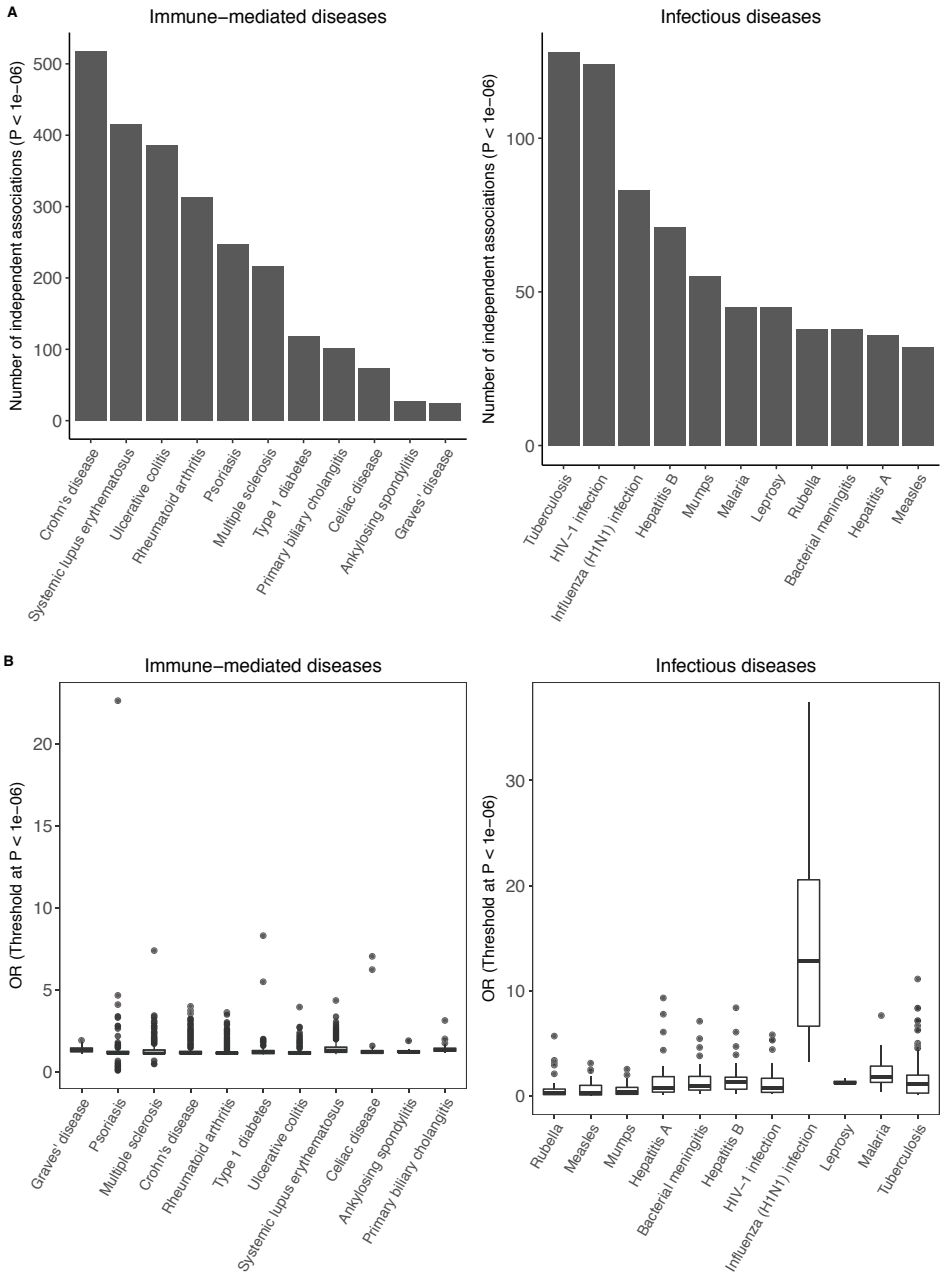



**Fig. 1.** (A) Number of genetic associations identified in genome-wide association studies (GWAS) with P value < $1x10^{-6}$, as recorded in the GWAS catalogue so far (https://www.ebi.ac.uk/gwas/), for a number of common immune-mediated and infectious diseases. (B) Effect size (odds ratio, OR) of immune-mediated diseases versus infectious diseases, extracted from the GWAS catalogue (https://www.ebi.ac.uk/gwas/).

**Box 1. Brief description of fungal pathogens**

***Candida albicans.*** Currently, there are 200 species in the genus *Candida*. These are collectively responsible for as many as 400,000 cases of systemic fungal diseases per year worldwide, with mortality rates of up to 40%[34–37]. The limited arsenal of anti-fungal drugs, which are classified into three classes, and the emergence of drug resistance through multiple mechanisms, including the formation of biofilms, make *Candida* the fourth most common cause of nosocomial infections[38–44]. Of these species, *Candida albicans* (*C. albicans*) is one of the best studied and most prevalent of the human fungal pathogens[45]. It is an opportunistic pathogen and, under physiological conditions, resides in the gut, genito-urinary tract and skin. Infection occurs only if the epithelial barrier function is impaired and/or there are microbiome imbalances and/or the host's immune system is compromised. It is generally believed that the main reservoir for *C. albicans* in humans is in the gastrointestinal tract and, once this barrier is disrupted, *C. albicans* can cause systemic infections[46,47].

Under pathogenic conditions, it can lead to superficial infections of the skin up to severe mucosal or life-threatening systemic infections, with the latter being associated with high mortality rates. One of the critical properties that determines the virulence of *C. albicans* is the phenotype switching between unicellular yeast cells and filamentous forms, called hyphae or pseudohyphae, which is known as morphological dimorphism (Fig. 2)[48,49]. This fungal dimorphism facilitates invasion of host tissues, escape from macrophages, dissemination in the bloodstream, and differentially affects immune recognition[50,51]. *C. albicans* is able to switch morphology in response to various environmental triggers, including a rise in temperature to 37°C, a pH equal to or higher than 7.0, a $CO_2$ concentration of 5.5% or the presence of serum or carbon sources, which stimulate hyphal growth[52,53]. Hyphae forms cause tissue damage and invasion, whereas yeast cells are thought to be responsible for dissemination to the environment and to new hosts[54].

***Aspergillus fumigatus.*** *Aspergillus fumigatus* (*A. fumigatus*) is the most prevalent airborne, filamentous, saprophytic and ubiquitous fungus responsible for mould. It is found worldwide and its conidia have a small diameter of 2–3 μm; they are abundant in the normal environment[55]. Everyone can inhale airborne *A. fumigatus* conidia, because their concentration in air is high (approx. 1–100 conidia per m3)[56]. Upon inhalation, the fungus penetrates alveolar epithelial and endothelial cells before an effective adaptive immune response is mounted against the fungus[57]. Depending on the host's immune status, this fungus can cause a range of pulmonary diseases, including allergic, invasive and non-invasive diseases. These diseases include *invasive pulmonary aspergillosis* (IPA), which is a severe disease characterized by tissue destruction, primarily affecting patients with severe immunodeficiency, but also critically ill patients without neutropenia and patients with chronic obstructive pulmonary disease (COPD)[58,59]. *Chronic pulmonary aspergillosis* includes aspergilloma and progressive destructive cavitary disease (also known as *chronic necrotizing pulmonary aspergillosis*)[58,59]. Aspergilloma is a fungus ball that usually develops in a patient's body cavity such as a paranasal sinus or an organ such as the lung; it is occasionally complicated by life-threatening haemoptysis. *Chronic necrotizing aspergillosis* is a locally invasive disease that mainly affects patients with chronic lung disease or mild immunodeficiency[59]. *Allergic bronchopulmonary aspergillosis* (ABPA) is a non-invasive form of lung disease that includes a hypersensitivity reaction to *Aspergillus* antigens; it is exclusively seen in patients with asthma or cystic fibrosis[59].

***Cryptococcus neoformans.*** *Cryptococcus neoformans* (*C. neoformans*) is a saprophytic, basidiomycetous, dimorphic organism found worldwide; its natural habitats are pigeon droppings and contaminated soil. Small-sized basidiospores (1.8 to 3.0 μm in diameter)[60,61] can form yeast cells, which are favoured at 37°C, or dikaryotic hyphae, favoured at 24°C[53]. When humans inhale basidiospores or yeast cells, the pathogen can disseminate throughout the respiratory tract leading to pulmonary infections. In some rare cases, *C. neoformans* can also infect the central nervous system leading to life-threatening meningoencephalitis, in both immuno-compromised and immuno-competent patients. If meningoencephalitis is left untreated, it is always fatal, while even after treatment the mortality rate lies between 10–25%[60,62].

approach, known as quantitative trait loci (QTL) mapping, genetic variants that regulate a given phenotype can be identified. In this way, we can connect phenotypes with genes and gene networks that better reflect the complex nature of host-pathogen interactions. For instance, genetic variants that influence gene or protein expression in close proximity to the gene of interest are known as *cis*-expression quantitative trait loci (eQTLs) or *cis*-protein quantitative trait loci (pQTLs). When genetic variants affect gene expression or protein levels in a distal manner, they are known as *trans*-eQTLs or *trans*-pQTLs. Importantly, it has been found that some diseases showed marked cell-type specificity, which means that genetic variants associated with a certain disease are enriched for cell-type-specific *cis*-eQTLs[22,23]. Furthermore, a systems genomics approach is a hypothesis-free approach, which can reveal novel, previously unknown associations. In general, studies of infectious diseases in humans are compromised because of patients' uncertain history of lifetime infections and the pathogens involved; moreover, experimental infection in volunteers is performed to a very limited extent for obvious ethical reasons. Therefore, another advantage of using a systems approach in infectious diseases is that we can generate and integrate molecular data from *in vitro* (cell lines, stem cells, etc.) and *in vivo* experimental models (animal models) and translate the results into a clinical situation by predicting disease outcome in human patients. For instance, in the case of influenza in which the responses in mice are very similar to those in human and non-human primates, the experimental animal models represent excellent systems for using systems biology to gain a comprehensive understanding of influenza-host interactions[24].

Thus, in the last few years, we have witnessed how the study of host-pathogen interactions during infection has moved towards implementing a systems approach. In this thesis, my aim is to identify risk genes and molecular pathways underlying the immune defence to infectious diseases by implementing a systems genomics approach. My main focus has been on fungal diseases, particularly those caused by the opportunistic fungal pathogen *Candida albicans* (*C. albicans*). Next, I describe what is currently known about genetics and immune defence mechanisms during *C. albicans* infections and provide an outline of the thesis chapters.

## • Why Do We Study Fungal Diseases?

Fungal diseases kill more than 1.5 million people worldwide every year and affect over a billion people[25]. In recent decades, the problem of severe nosocomial fungal diseases has become more serious, especially in immune-compromised patients. The risk factors for developing fungal diseases have been well-described[26], although not all at-risk patients will develop disease, suggesting that host genetics also determines susceptibility to fungal infections. Although the epidemiology of fungal diseases has changed greatly over the past few decades, the most frequently diagnosed fungal infections are caused by

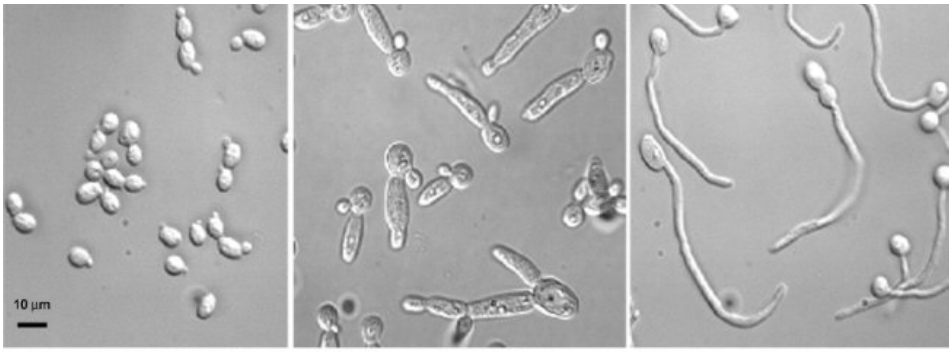



**Fig. 2.** Forms of *Candida albicans*. *C. albicans* is referred to as dimorphic fungus since it grows as round yeast cells and as filamentous cells known as hyphae of pseudo hyphae, depending on the micro-environment. Picture adapted from: http://www.usask.ca/biology/kaminskyj/images/BachewichCandida.jpg, University of Saskatchewan, Canada

pathogens from the genera *Candida*, *Cryptococcus* and *Aspergillus*[27] (Box 1). These fungi are ubiquitous and, in particular, *Cryptococcus neoformans* (*C. neoformans*) and *Aspergillus fumigatus* (*A. fumigatus*) can be found in the host environment, whereas *Candida albicans* (*C. albicans*) is part of the normal microbiota of an individual's skin, mucocutaneous surfaces and gastro-intestinal tract[28]. *C. albicans* is the main cause of mucosal and systemic infections, *A. fumigatus* of most allergic fungal diseases, and *C. neoformans* of lung diseases that may spread to the brain and cause meningoencephalitis. Diagnosing a fungal infection is often difficult due to the presence of non-specific symptoms, while challenges in isolating and identifying fungi make it ineffective to prevent and treat these infections[29,30]. There are no vaccines available and, despite the availability of potent antifungal drugs, fungal diseases have a high mortality, estimated at 1.5 million annually[25]. In addition, identifying novel drug targets and developing effective antifungal drugs is complicated by the similarity between eukaryotic fungi and humans.

*C. albicans* is considered to be the major fungal pathogen of humans in the heterogeneous group of *Candida* species (Box 1, Fig. 2)[31]. It is an opportunistic fungus that is potentially pathogenic when the immune system is weakened causing superficial or systemic infections (candidaemia). Several risk factors are known to increase the risk of developing *Candida* infections, including a prolonged stay in intensive care (ICU), use of immunosuppressive agents/antibiotics, transplantation, surgical intervention, neutropenia, solid tumour and haematological malignancies, and parenteral nutrition[32]. Although these factors are important, they do not explain all *Candida* infections and only a minority of patients at risk will actually develop disease, suggesting the critical role of genetics in determining

disease susceptibility. Indeed, several monogenetic disorders that result in severe primary immunodeficiencies, and mutations and common polymorphisms in immune genes, have been associated with an increased susceptibility to mucosal and/or invasive *Candida* infections (Box 2).

**Box 2. Genetic susceptibility to *Candida* infection.** Several monogenetic disorders have been linked to increased susceptibility to both mucosal and systemic *Candida* infections. These disorders have mutations in immune genes, such as *CARD9*, *STAT1*, *IL-23* and *IL-22* as heterodimer with *STAT3* or *STAT4*, *STAT3*, *IL-17A*, *IL-17F*, dedicator of cytokinesis (*DOCK*)*8* and *TYK2*; they have been reviewed in detail elsewhere[63]. Patients carrying mutations in genes coding for cytokines and their receptors, such as *IL-12RB1* and *CD25*, also have an increased susceptibility to *Candida* infections[64]. Moreover, an increased susceptibility to *Candida* (both mucosal and systemic infections) can also be explained by the presence of common polymorphisms mainly in genes that encode cytokines and receptors that recognize fungal antigens, such as *Dectin-1*, *TLR-1*, *TLR-2*, *TLR3*, *TLR-4*, *MBL-2*, *IL-4*, *IL-10*, *IL-12b*, *DEFB1*, (coding for beta-defensin 1), *PTPN22* and *NLPR3*[63]. Lastly, a genome-wide screen of approximately 200,000 SNPs in 186 loci recently identified three more genes (*CD58*, *LCE4A-C1orf68* and *TAGAP*) associated with candidaemia susceptibility [65].

Despite significant progress over the last few years in identifying susceptibility genes for *Candida* infections, there is still much genetic information unexplored and particularly the molecular mechanisms underlying susceptibility are not fully understood. Both innate and adaptive immune responses are important for host immune defence against fungal pathogens (Box 3).

However, it is difficult to identify genetic factors using a traditional GWAS because of the limited size of patient cohorts and use of inappropriate controls (such as individuals with asymptomatic infections). In the case of invasive *Candida* infections, it is difficult to gather large cohorts of patients, as they are relatively scarce, especially in individual medical centres. In addition, a GWAS alone cannot explain how genetic variation affects disease or which organ or tissue, or even which cell type, will be affected. To overcome these challenges, we use a systems genomics approach to identify risk loci and molecular pathways underlying host immune defence and disease pathogenesis. In particular, to investigate the effect of *Candida* infection on host immune response, we explore changes at the transcriptomic or proteomic level in a relevant mixture of lymphocytes (T cells, B cells, and NK cells) and monocytes, i.e. in peripheral blood mononuclear cells (PBMCs), which are easily available. By integrating different molecular data (genomics, transcriptomics, proteomics and immunological data), we aim to gain a better understanding of the host immune defence mechanisms against *C. albicans* (Fig. 3). In the next section, I present an outline of this thesis.

**Box 3. Immune defence against *Candida* fungal infections.** The innate immune system forms the first line of defence against all invading organisms. Its first response involves recognizing invading fungi from immune cells, such as phagocytes, via specialized receptors that recognize conserved pathogen-associated molecular patterns (PAMPs), which are known as pattern recognition receptors (PRRs)[66–68]. This recognition induces signalling-mediated gene transcription that leads to the secretion of inflammatory proteins, such as chemokines and cytokines[66–68]. In turn, the inflammatory proteins recruit neutrophils and other immune cells to the infection site, killing the fungal cells and activating the adaptive immune responses[66–68].

Neutrophils play a pivotal role in controlling fungal infections, with neutropenia being a major risk factor for invasive fungal infections[69,70]. These cells kill *Candida* cells through phagocytosis, oxidative and non-oxidative effector mechanisms, and the release of neutrophil extracellular traps (NETs)[71–73]. The PRRs include the well-known Toll-like receptors (TLRs; TLR2 and 4), C-type lectin receptors (CLRs, such as dectin-1, dectin-2 and mannose receptor (MR)), NOD-like receptors (NLRs), and RIG-I-like receptors (RLRs)[74].

The fungal PAMPs mainly consist of polysaccharides that are present on the fungal cell wall. Of note, the polysaccharide structures differ between yeasts and hyphae of *C. albicans*, which can explain the differential immune responses reported between the two forms of *C. albicans*[50,75–81]. As a crucial part of the innate immune response, early studies showed that *C. albicans* activates the complement system via all three pathways (the classical, the alternative and the lectin pathway)[82–85]. The interaction of *C. albicans* with the complement system has been reviewed in detail [86]. Activation of the complement system by *C. albicans* contributes to the enhanced induction of pro-inflammatory cytokines in PBMCs, such as IL-6 and IL-1β, by the activation of the anaphylatoxin C5a[87]. In an attempt to escape the immune surveillance, *C. albicans* has been shown to bind two central complement regulator factors, H and FHL-1, from human serum and it thus escapes the control of the complement system[88]. Platelets are less appreciated players of the host immune defence, despite the fact that the interaction of *C. albicans* with platelets has been reported in early *in vitro* and *in vivo* studies[89–91]. Platelet-rich plasma presents anti-microbial peptides with anti-candida activity, such as the chemokine RANTES, platelet-factor-4 (PF-4) and thrombocidin-1 (TC-1), and it also inhibits candida growth[92–94].

## • Outline of the Thesis

In *chapter 2*, we demonstrate the potential of using a systems genomics approach to identify genes and pathways underlying susceptibility to *Candida* infection. In this study, we integrated genetic data from our candidaemia patient cohort with gene-expression profiles recorded upon *Candida* stimulation in PBMCs isolated from eight healthy volunteers, *Candida*-induced cytokine production capacity in PBMCs (TNFα, IL-6 and IFNγ), and circulating concentrations of three pro-inflammatory cytokines (IL-6, IFNγ and IL-8) in patient serum. Our candidaemia cohort consisted of 217 patients of European ancestry and genotyped with the Immunochip SNP array that covers 186 loci[33]. For association testing, we followed a two-stage analysis and used two different control groups: 11,920 population-based healthy controls in the first stage and 146 case-matched but candidiasis-free controls in the second stage (Table 2). By integrating all these different datasets, we could prioritize genes from the Immunochip study and investigate the pathways in which these genes are enriched. Lastly, we followed up a gene from one of the

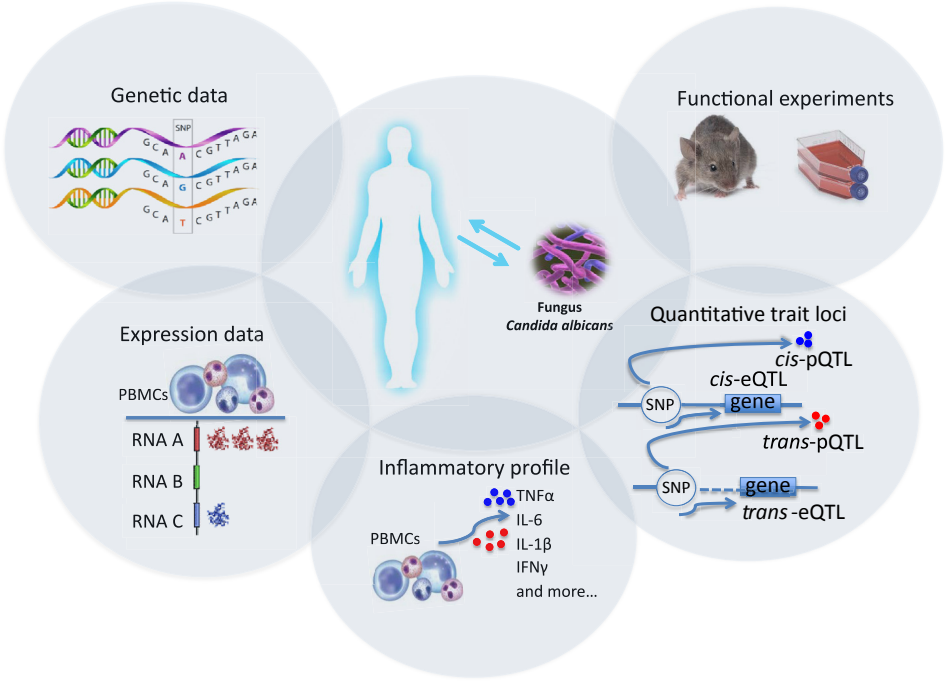



**Fig. 3.** A systems genomics approach towards understanding the interactions of *C. albicans* with its human host. To identify risk loci and pathways to *Candida i*nfection, we integrate different molecular data (from left to right): (1) Genome-wide genetic data generated from our candidaemia cohort that was used for the GWAS. (2) To study how the gene immune responses are modulated by *Candida* infection, we used RNA sequencing to generate expression data from PBMCs stimulated by *C. albicans*. (3) To further investigate response modulation by *Candida* infection, we measured the inflammatory responses from PBMCs stimulated by *C. albicans*. By using genetic data and expression data or measurements of inflammatory proteins, we were able to map the genetic variation that influences gene expression or inflammatory proteins, respectively, i.e. expression-quantitative trait loci (eQTLs) or protein-QTLs (pQTLs). Our ultimate goal is to prioritize good candidates genes for follow-up *in vitro* and *in vivo* functional experiments, leading to a better understanding of their role in human immune defence against *C. albicans*.

candidaemia-associated loci (*MAP3K8*) with *in vitro* experiments to validate its role in cytokine regulation in response to *Candida* infection.

In *chapter 3*, we first describe exploring the genetic architecture of cytokine responses induced by the two forms of *C. albicans* (yeast and hyphae). Then we extended our studies to the two other common fungal pathogens acquired from the environment (*A. fumigatus* and *C. neoformans*) and investigated the hypothesis that susceptibility to candidaemia is explained by modulating the levels of pro-inflammatory cytokines. To identify genetic loci that regulate cytokine responses in response to stimulation with these three opportunistic fungi, we made use of genetic data and measurements of six cytokines (IL-6, TNFα, IL-1β, IL-17,

IL-22 and IFNγ) from a large, population-based cohort of European ancestry, the 500 Human Functional Genomics (500FG) cohort (Human Functional Genomics Project, HFGP, see www.humanfunctionalgenomics.org). Using these data, we mapped genetic variation that influences cytokine levels, i.e. cytokine-quantitative trait loci (cQTLs). cQTL studies can offer mechanistic insight into how genetic variation can affect disease and, thus, help in prioritizing genes that potentially cause disease. The measured cytokines were derived from three different cell systems (PBMCs, whole blood and monocyte-derived macrophages) stimulated by the three fungi (Table 2). Since the Immunochip SNP array covers only 3% of the genome, we went on to perform the first GWAS to date to identify novel susceptibility loci in our candidaemia patient cohort (Table 2) and to explore whether genetic variants that influence cytokine levels are also associated with susceptibility to candidaemia.

With the aim of systematically exploring the inflammatory proteins released into the blood circulation upon *Candida* stimulation, we extended our studies in *chapter 4* by profiling 92 inflammatory proteins in plasma from our candidaemia patient cohort (n = 42) and from healthy individuals in the Lifelines cohort (https://www.lifelines.nl/researcher/biobank-lifelines/) (n = 89), and also in *Candida*-stimulated PBMCs from two independent population-based cohorts, the 500FG (n = 360) and Lifelines DEEP cohort (https://www.lifelines.nl/researcher/biobank-lifelines/additional-studies/lifelines-deep) (n = 76), using the high-throughput technology OLINK (https://www.olink.com/). Using genetic data and protein measurements from the 500FG and Lifelines DEEP cohorts, we investigated the genetic factors that determine inflammatory responses in the context of *Candida* infection by performing genome-wide protein-quantitative-trait loci (pQTL) mapping. We also checked whether genetic variation influencing inflammatory responses could explain susceptibility to candidaemia and survival outcome among candidaemia patients. Table 2 describes the cohorts, technologies and molecular data that we generated in the context of *Candida* infection for this PhD research.

In *chapter 5*, we focus on the major histocompatibility complex (MHC) locus, also known as the human leukocyte antigen (HLA), as it plays an important role in recognition of pathogenic antigens during host immune defence and it has been reported as a major risk factor for complex diseases. We review recent advances that have contributed to a better understanding of the role of MHC in two major classes of complex diseases: autoimmune and infectious diseases. Of note, we compare the associations reported so far for autoimmune and infectious diseases and discuss the role of pathogens in autoimmune disease development.

Lastly, in *chapter 6*, we discuss the challenges of studying the genetics of infectious diseases, what we have learned from using a systems genomics approach in *Can-*

*dida* infection and the limitations of our work. We also describe our findings in a broader perspective and future directions for studying the genetics of infectious diseases. In summary, we aimed to implement a systems genomics approach to reveal the genetic architecture and molecular pathways that underlie susceptibility to candidaemia. By integrating different layers of molecular data, our ultimate goal was to pinpoint and prioritize genes that could serve as potential immunotherapeutic targets to improve patient outcome (Fig. 3). Importantly, we aimed to provide a framework to identify genes and molecular pathways not only for candidaemia, but also for other infectious diseases.

**Table 2.** Summary of the cohorts and molecular data used in this thesis.

**Genetic data**

| Candidaemia cohort | Number of individuals | Genotyping array | Imputation | Chapter |
|---|---|---|---|---|
| ***Immunochip-wide association analysis*** | | | | |
| Cases | 217 (European ancestry) | Immunochip (Illumina) | no | 2 |
| Controls (population-based) | 11,920 (European ancestry) | Immunochip (Illumina) | no | 2 |
| Controls (case-matched) | 146 (European ancestry) | Immunochip (Illumina) | no | 2 |
| ***GWAS*** | | | | |
| Cases | 178 (Europeans) of which | HumanCoreExome-24 v1.0 and | yes (HRC*) | 3 |
| | 31 non-surv.** ( Days < 14) | HumanCoreExome-12 v1.0 (Illumina) | | 4 |
| | 18 non-surv.** ( 14 < Days < 30 ) | | | 4 |
| | 16 non-surv.** ( 30 < Days < 90 ) | | | 4 |
| | 83 survivors | | | 4 |
| Controls (case-matched) | 175 (European ancestry) | HumanCoreExome-24 v1.0 and HumanCoreExome-12 v1.0(Illumina) | yes (HRC*) | 3 |
| ***Population-based cohorts*** | | | | |
| 200FG | ~200 healthy individuals (Europeans) | ImmunoChip (Illumina) | yes (HRC*) | 3 |
| 500FG | 534 healthy individuals (Europeans) | Illumina Human OmniExpress Exome-8 v1.0 | yes (GoNL***) | 3, 4 |
| Lifelines Deep | 1539 healthy individuals (Europeans) | CytoSNP, ImmunoChip (Illumina) | yes (GoNL***) | 4 |

**Transcriptomic data**

| Cohort | Cell system | Stimulation time | Stimulant | Technique | Chapter |
|---|---|---|---|---|---|
| 8 healthy volunteers | PBMCs | 4 and 24 hours | *C. albicans* yeast | RNA sequencing | 2 |

**Cytokine data**

| Cohort | Cell system | Cytokines | Stimulation time | Stimulant | Technique | Chapter |
|---|---|---|---|---|---|---|
| Healthy volunteers | PBMCs | TNFα, IL-6 | 1 day | *C. albicans* yeast | ELISA | 2 |
| Healthy volunteers | PBMCs | IFNγ | 2 days | *C. albicans* yeast | ELISA | 2 |
| Candidaemia patients | serum | IL-6, IL-8, IFNγ | | | ELISA | 2 |
| 500FG | macrophages | IL-6, TNFα | 1 day | *C. albicans* yeast | ELISA | 3 |
| 500FG | PBMCs | IL-6, TNFα, IL-1β | 1 day | *C. albicans* yeast | ELISA | 3 |
| 500FG | blood | IL-6, TNFα, IL-1 β,IFNγ | 2 days | *C. albicans* yeast | ELISA | 3 |
| 500FG | PBMCs | IL-17, IL-22, IFNγ | 7 days | *C. albicans* yeast | ELISA | 3 |
| 500FG | PBMCs | IL-6, TNFα, IL-1 β | 1 day | *C. albicans* hyphae | ELISA | 3 |
| 500FG | PBMCs | IL-17, IL-22, IFNγ | 7 days | *C. albicans* hyphae | ELISA | 3 |
| 500FG | PBMCs | IL-6, TNFα | 1 day | *A. fumigatus* | ELISA | 3 |
| 500FG | PBMCs | IL-17, IL-22, IFNγ | 7 days | *A. fumigatus* | ELISA | 3 |
| 500FG | PBMCs | IL-6, TNFα, IL-1 β | 1 day | *C. neoformans* | ELISA | 3 |
| 500FG | PBMCs | IL-17, IL-22, IFNγ | 7 days | *C. neoformans* | ELISA | 3 |
| 500FG (n = 360) | PBMCs | 92 inflammatory proteins | 1 day | *C. albicans* yeast | OLINK | 4 |
| Lifelines Deep (n= 76) | PBMCs | 92 inflammatory proteins | 1 day | *C. albicans* yeast | OLINK | 4 |
| Lifelines (n = 89) | plasma | 92 inflammatory proteins | | | OLINK | 4 |
| Cases (n = 42) | plasma | 92 inflammatory proteins | | | OLINK | 4 |

*HRC Human Reference Consortium, used as reference panel for imputation, **non-surv. = non-survivors, ***GoNL Genome of the Netherlands, used as a reference panel for imputation, , PBMCs peripheral blood mononuclear cells

# References

1. Black, R. E. et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 375, 1969–1987 (2010).
2. Dye, C. After 2015: infectious diseases in a new era of health and development. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 369, 20130426 (2014).
3. Fantus, B. The therapy of the Cook County Hospital. J. Am. Med. Assoc. 111, 317–321 (1938).
4. Allison, A. C. Protection afforded by sickle-cell trait against subtertian malarial infection. Br. Med. J. 1, 290–294 (1954).
5. Misch, E. A., Berrington, W. R., Vary, J. C. & Hawn, T. R. Leprosy and the human genome. Microbiol. Mol. Biol. Rev. 74, 589–620 (2010).
6. Herndon, C.N., Jennings, R. G. A twin-family study of susceptibility to poliomyelitis. Am. J. Hum. Genet. 3, 17–46 (1951).
7. Comstock, G. W. Tuberculosis in twins: a re-analysis of the Prophit survey. Am Rev Respir Dis 117, 621–624 (1978).
8. Kallmann, F.J., Reisner, D. Twin Studies on the Significance of Genetic Factors in Tuberculosis. Am. Rev. Tuberc. Pulm. Dis. 47, 549–71 (1943).
9. Van Der Eijk, E. A., Van De Vosse, E., Vandenbroucke, J. P. & Van Dissel, J. T. Heredity versus environment in tuberculosis in twins: The 1950s United Kingdom Prophit Survey - Simonds and Comstock revisited. Am. J. Respir. Crit. Care Med. 176, 1281–1288 (2007).
10. Malaty, H.M., Engstrand, L., Pedersen, N.L., Graham, D. Y. Helicobacter pylori infection: genetic and environmental influences. A study of twins. Ann. Intern. Med. 120, 982–986 (1994).
11. Lin, T. M. et al. Hepatitis B virus markers in Chinese twins. Anticancer Res 9, 737–741 (1989).
12. Sørensen, T. I. A., Nielsen, G. G., Andersen, P. K. & Teasdale, T. W. Genetic and environmental influences on premature death in adult adoptees. N. Engl. J. Med. 318, 727–732 (1988).
13. Marquet, S. et al. Genetic localization of a locus controlling the intensity of infection by Schistosoma mansoni on chromosome 5q31-q33. Nat. Genet. 14, 181–184 (1996).
14. Thye, T., Burchard, G. D., Nilius, M., Müller-Myhsok, B. & Horstmann, R. D. Genomewide linkage analysis identifies polymorphism in the human interferon-gamma receptor affecting Helicobacter pylori infection. Am. J. Hum. Genet. 72, 448–453 (2003).
15. Mira, M. T. et al. Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. Nat. Genet. 33, 412–415 (2003).
16. Siddiqui, M. R. et al. A major susceptibility locus for leprosy in India maps to chromosome 10p13. Nat. Genet. 27, 439–441 (2001).
17. Mira, M. T. et al. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 427, 636–640 (2004).
18. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–61 (2007).
19. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 526, 68–74 (2015).
20. Laubenbacher, R. et al. A systems biology view of cancer. Biochimica et Biophysica Acta - Reviews on Cancer 1796, 129–139 (2009).
21. Aderem, A. et al. A systems biology approach to infectious disease research: Innovating the pathogen-host research paradigm. MBio 2:e00325-10, (2011).
22. Gregersen, P. K. Cell type-specific eQTLs in the human immune system. Nature Genetics 44, 478–480 (2012).
23. Bahcall, O. Immune cell–specific eQTLs. Nat. Genet. 46, 532 (2014).
24. Kollmus, H., Wilk, E. & Schughart, K. Systems biology and systems genetics-novel innovative approaches to study host-pathogen interactions during influenza infection. Curr. Opin. Virol. 6C, 47–54 (2014).
25. Bongomin, F., Gago, S., Oladele, R. & Denning, D. Global and Multi-National Prevalence of Fungal Diseases—Estimate Precision. J. Fungi 3, 57 (2017).
26. Perfect, J. & Casadevall, A. in Molecular Principles of Fungal Pathogenesis. (ed. Heitman, J., Filler, S., Edwards, Jr. J, M. A.) 3–11 (ASM Press, Washington, DC) 3-11 (2006). doi:10.1128/9781555815776.ch1
27. Richardson, M. D. Changing patterns and

trends in systemic fungal infections. J Antimicrobial Chemotherapy 56, Suppl 1:i5-i11 (2005).
28. Underhill, D. M. & Iliev, I. D. The mycobiota: interactions between commensal fungi and the host immune system. Nat. Rev. Immunol. 14, 405–16 (2014).
29. Hinrikson, H. P., Hurst, S. F., De Aguirre, L. & Morrison, C. J. Molecular methods for the identification of Aspergillus species. Med. Mycol. 43 Suppl 1, S129–S137 (2005).
30. Caliendo, A. M. et al. Better Tests, better care: improved diagnostics for infectious diseases. Clin. Infect. Dis. 57, S139–S170 (2013).
31. Turner, S. A. & Butler, G. The Candida pathogenic species complex. Cold Spring Harb. Perspect. Med. 4(9):a019778 (2014).
32. Yapar, N. Epidemiology and risk factors for invasive candidiasis. Therapeutics and Clinical Risk Management 10, 95–105 (2014).
33. Cortes, A. & Brown, M. A. Promise and pitfalls of the Immunochip. Arthritis Res. Ther. 13, 101 (2011).
34. Gudlaugsson, O. et al. Attributable mortality of nosocomial candidemia, revisited. Clin. Infect. Dis. 37, 1172–1177 (2003).
35. Pappas, P. G. et al. A prospective observational study of candidemia: epidemiology, therapy, and influences on mortality in hospitalized adult and pediatric patients. Clin. Infect. Dis. 37, 634–643 (2003).
36. Campion, E. W., Kullberg, B. J. & Arendrup, M. C. Invasive Candidiasis. N. Engl. J. Med. 373, 1445–1456 (2015).
37. Barnett, J. A. A history of research on yeasts 8: Taxonomy. Yeast 21, 1141–1193 (2004).
38. Prasad, R., Shah, A. H. & Rawal, M. K. Antifungals: Mechanism of action and drug resistance. Adv. Exp. Med. Biol. 892, 327–349 (2016).
39. Wisplinghoff, H. et al. Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. Clin. Infect. Dis. 39, 309–17 (2004).
40. Wisplinghoff, H. et al. Nosocomial bloodstream infections in pediatric patients in United States hospitals: epidemiology, clinical features and susceptibilities. Pediatr. Infect. Dis. J. 22, 686–691 (2003).
41. Wisplinghoff, H., Seifert, H., Wenzel, R. P. & Edmond, M. B. Current trends in the epidemiology of nosocomial bloodstream infections in patients with hematological malignancies and solid neoplasms in hospitals in the United States. Clin. Infect. Dis. 36, 1103–1110 (2003).
42. Douglas, L. J. Candida biofilms and their role in infection. Trends in Microbiology 11, 30–36 (2003).
43. Tournu, H. & Van Dijck, P. Candida biofilms and the host: Models and new concepts for eradication. International J Microbiology (2012). doi:10.1155/2012/845352
44. Taff, H. T. et al. A Candida Biofilm-Induced Pathway for Matrix Glucan Delivery: Implications for Drug Resistance. PLoS Pathog. 8(8):e1002848 (2012).
45. Bassetti, M. et al. Epidemiology, species distribution, antifungal susceptibility and outcome of nosocomial candidemia in a tertiary care hospital in Italy. PLoS One 6(9):e24198 (2011).
46. Schulze, J. & Sonnenborn, U. Yeasts in the gut: from commensals to infectious agents. Dtsch. Arztebl. Int. 106, 837–42 (2009).
47. Nucci, M. & Anaissie, E. Revisiting the source of candidemia: skin or gut? Clin. Infect. Dis. 33, 1959–1967 (2001).
48. Lo, H. J. et al. Nonfilamentous C. albicans mutants are avirulent. Cell 90, 939–949 (1997).
49. Sudbery, P., Gow, N. & Berman, J. The distinct morphogenic states of Candida albicans. Trends in Microbiology 12, 317–324 (2004).
50. Mukaremera, L., Lee, K. K., Mora-Montes, H. M. & Gow, N. A. R. Candida albicans yeast, pseudohyphal, and hyphal morphogenesis differentially affects immune recognition. Front. Immunol. 8, 629 (2017). doi: 10.3389/fimmu.2017.00629
51. McKenzie, C. G. J. et al. Contribution of Candida albicans cell wall components to recognition by and escape from murine macrophages. Infect. Immun. 78, 1650–1658 (2010).
52. Hall, R. A. et al. CO2 acts as a signalling molecule in populations of the fungal pathogen Candida albicans. PLoS Pathog. 6(11):e1001193 (2010).
53. Karkowska-Kuleta, J., Rapala-Kozik, M. & Kozik, A. Fungi pathogenic to humans: Molecular bases of virulence of Candida albicans, Cryptococcus neoformans and Aspergillus fumigatus. Acta Biochimica Polonica 56, 211–224 (2009).

54. Sudbery, P. E. Growth of Candida albicans hyphae. Nat Rev Microbiol 9, 737–748 (2011).
55. Latgé, J. P. Aspergillus fumigatus and Aspergillosis. Clin. Microbiol. Rev. 12, 310–350 (1999).
56. Latgé, J. P. The pathobiology of Aspergillus fumigatus. Trends in Microbiology 9, 382–389 (2001).
57. Hope, W. W. et al. Pathogenesis of Aspergillus fumigatus and the kinetics of galactomannan in an in vitro model of early invasive pulmonary aspergillosis: implications for antifungal therapy. J. Infect. Dis. 195, 455–466 (2007).
58. Kousha, M., Tadi, R. & Soubani, A. O. Pulmonary aspergillosis: A clinical review. European Respiratory Review 20, 156–174 (2011).
59. Patterson, K. C. & Strek, M. E. Diagnosis and treatment of pulmonary aspergillosis syndromes. Chest 146, 1358–1368 (2014).
60. Buchanan, K. L. & Murphy, J. W. What makes Cryptococcus neoformans a pathogen? Emerging Infectious Diseases 4, 71–83 (1998).
61. Lin, X. & Heitman, J. Chlamydospore formation during hyphal growth in Cryptococcus neoformans. Eukaryot. Cell 4, 1746–1754 (2005).
62. Perfect, J. R. & Casadevall, A. Crypto coccosis. Infectious Disease Clinics of North America 16, 837–874 (2002).
63. Smeekens, S. P., van de Veerdonk, F. L., Kullberg, B. J. & Netea, M. G. Genetic susceptibility to Candida infections. EMBO Mol. Med. 5, 805–813 (2013).
64. Smeekens, S. P. et al. Autophagy is redundant for the host defense against systemic Candida albicans infections. Eur. J. Clin. Microbiol. Infect. Dis. 33, 711–722 (2014).
65. Kumar, V. et al. Immunochip SNP array identifies novel genetic variants conferring susceptibility to candidaemia. Nat. Commun. 5, 4675 (2014). doi: 10.1038/ncomms5675
66. Netea, M. G., Brown, G. D., Kullberg, B. J. & Gow, N. a R. An integrated model of the recognition of Candida albicans by the innate immune system. Nat. Rev. Microbiol. 6, 67–78 (2008).
67. Hoebe, K., Janssen, E. & Beutler, B. The interface between innate and adaptive immunity. Nature Immunology 5, 971–974 (2004).
68. Medzhitov, R. Recognition of microorganisms and activation of the immune response. Nature 449, 819–826 (2007).
69. Horn, D. L. et al. Epidemiology and Outcomes of Candidemia in 2019 Patients: data from the prospective antifungal therapy alliance registry. Clin. Infect. Dis. 48, 1695–1703 (2009).
70. Uzun, O., Ascioglu, S., Anaissie, E. J. & Rex, J. H. Risk factors and predictors of outcome in patients with cancer and breakthrough candidemia. Clin. Infect. Dis. 32, 1713–7 (2001).
71. Vonk, A. G., Wieland, C. W., Netea, M. G. & Kullberg, B. J. Phagocytosis and intracellular killing of Candida albicans blastoconidia by neutrophils and macrophages: A comparison of different microbiological test systems. J. Microbiol. Methods 49, 55–62 (2002).
72. Aratani, Y. et al. Critical role of myeloperoxidase and nicotinamide adenine dinucleotide phosphate-oxidase in high-burden systemic infection of mice with Candida albicans. J. Infect. Dis. 185, 1833–1837 (2002).
73. Urban, C. F., Reichard, U., Brinkmann, V. & Zychlinsky, A. Neutrophil extracellular traps capture and kill Candida albicans and hyphal forms. Cell. Microbiol. 8, 668–676 (2006).
74. Netea, M. G., Joosten, L. A., van der Meer, J. W., Kullberg, B. J. & van de Veerdonk, F. L. Immune defence against Candida fungal infections. Nat Rev Immunol 15, 630–642 (2015).
75. Shibata, N., Suzuki, A., Kobayashi, H. & Okawa, Y. Chemical structure of the cell-wall mannan of Candida albicans serotype A and its difference in yeast and hyphal forms. Biochem. J. 404, 365–372 (2007).
76. Lowman, D. W. et al. Novel structural features in Candida albicans hyphal glucan provide a basis for differential innate immune recognition of hyphae versus yeast. J. Biol. Chem. 289, 3432–3443 (2014).
77. Munro, C. A., Schofield, D. A., Gooday, G. W. & Gow, N. A. R. Regulation of chitin synthesis during dimorphic growth of Candida albicans. Microbiology 144, 391–401 (1998).
78. Gow, N. A., van de Veerdonk, F. L., Brown, A. J. & Netea, M. G. Candida albicans morphogenesis and host defence: discriminating invasion from colonization. Nat. Rev. Microbiol. 10, 112–122 (2011).
79. Gantner, B. N., Simmons, R. M. & Underhill, D. M. Dectin-1 mediates macrophage

recognition of Candida albicans yeast but not filaments. EMBO J. 24, 1277–1286 (2005).
80. van der Graaf, C. A., Netea, M. G., Verschueren, I., van der Meer, J. W. & Kullberg, B. J. Differential cytokine production and toll-like receptor signaling pathways by Candida albicans blastoconidia and hyphae. Infect. Immun. 73, 7458–7464 (2005).
81. Moyes, D. L. et al. Candida albicans yeast and hyphae are discriminated by MAPK signaling in vaginal epithelial cells. PLoS One 6(11):e26580 (2011). doi: 10.1371/journal.pone.0026580
82. Zhang, M. X., Lupan, D. M. & Kozel, T. R. Mannan-specific immunoglobulin G antibodies in normal human serum mediate classical pathway initiation of C3 binding to Candida albicans. Infect. Immun. 65, 3822–3827 (1997).
83. Zhang, M. X. & Kozel, T. R. Mannan-specific immunoglobulin G antibodies in normal human serum accelerate binding of C3 to Candida albicans via the alternative complement pathway. Infect. Immun. 66, 4845–4850 (1998).
84. Sealy, P. I., Garner, B., Swiatlo, E., Chapman, S. W. & Cleary, J. D. The interaction of mannose binding lectin (MBL) with mannose containing glycopeptides and the resultant potential impact on invasive fungal infection. Med. Mycol. 46, 531–539 (2008).
85. Kozel, T. R. Activation of the complement system by pathogenic fungi. Clin. Microbiol. Rev. 9, 34–46 (1996).
86. Luo, S., Skerka, C., Kurzai, O. & Zipfel, P. F. Complement and innate immune evasion strategies of the human pathogenic fungus Candida albicans. Molecular Immunology 56, 161–169 (2013).
87. Cheng, S. C. et al. Complement plays a central role in Candida albicans-induced cytokine production by human PBMCs. Eur. J. Immunol. 42, 993–1004 (2012).
88. Meri, T. et al. The yeast Candida albicans binds complement regulators factor H and FHL-1. Infect. Immun. 70, 5185–5192 (2002).
89. Skerl, K. G., Calderone, R. A. & Sreevalsan, T. Platelet interactions with Candida albicans. Infect. Immun. 34, 938–943 (1981).
90. Robert, R. et al. Adherence of platelets to Candida species in vivo. Infect. Immun. 68, 570–576 (2000).
91. Willcox, M. D. P., Webb, B. C., Thakur, A. & Harty, D. W. S. Interactions between Candida species and platelets. J. Med. Microbiol. 47, 103–110 (1998).
92. Drago, L., Bortolin, M., Vassena, C., Taschieri, S. & Del Fabbro, M. Antimicrobial activity of pure platelet-rich plasma against microorganisms isolated from oral cavity. BMC Microbiol. 13, (2013).
93. Kwakman, P. H. S. et al. Native thrombocidin-1 and unfolded thrombocidin-1 exert antimicrobial activity via distinct structural elements. J. Biol. Chem. 286, 43506–43514 (2011).
94. Tang, Y. Q., Yeaman, M. R. & Selsted, M. E. Antimicrobial peptides from human platelets. Infect. Immun. 70, 6524–33 (2002).

# CHAPTER 2

## An integrative genomics approach identifies novel pathways that influence candidaemia susceptibility

Vasiliki Matzaraki, Mark S. Gresnigt, Martin Jaeger, Isis Ricaño-Ponce, Melissa D. Johnson, Marije Oosting, Lude Franke, Sebo Withoff, John R. Perfect, Leo A. B. Joosten, Bart Jan Kullberg, Frank L. van de Veerdonk, Iris Jonkers, Yang Li, Cisca Wijmenga, Mihai G. Netea, Vinod Kumar

PLOS ONE

RESEARCH ARTICLE

# An integrative genomics approach identifies novel pathways that influence candidaemia susceptibility

**Vasiliki Matzaraki[1], Mark S. Gresnigt[2], Martin Jaeger[2], Isis Ricaño-Ponce[1], Melissa D. Johnson[3], Marije Oosting[2], Lude Franke[1], Sebo Withoff[1], John R. Perfect[3], Leo A. B. Joosten[2], Bart Jan Kullberg[2], Frank L. van de Veerdonk[2], Iris Jonkers[1], Yang Li[1], Cisca Wijmenga[1,4], Mihai G. Netea[2,5☯], Vinod Kumar[1☯]***

**1** Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands, **2** Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, The Netherlands, **3** Duke University Medical Center, Durham, North Carolina, United States of America, **4** K.G. Jebsen Coeliac Disease Research Centre, Department of Immunology, University of Oslo, Oslo, Norway, **5** Human Genomics Laboratory, Craiova University of Medicine and Pharmacy, Craiova, Romania

☯ These authors contributed equally to this work.
* v.kumar@umcg.nl



OPEN ACCESS

**Citation:** Matzaraki V, Gresnigt MS, Jaeger M, Ricaño-Ponce I, Johnson MD, Oosting M, et al. (2017) An integrative genomics approach identifies novel pathways that influence candidaemia susceptibility. PLoS ONE 12(7): e0180824. https://doi.org/10.1371/journal.pone.0180824

**Editor:** Joy Sturtevant, Louisiana State University, UNITED STATES

**Received:** April 1, 2017

**Accepted:** June 21, 2017

**Published:** July 20, 2017



**Data Availability Statement:** All relevant data used within the paper are provided in the supplementary files.

**Funding:** This work was supported by an ERC Consolidator Grant [FP/2007-2013/ERC grant 2012-310372 to M.G.N.], an ERC Advanced grant [FP/2007-2013/ERC grant 2012-322698 to C.W.], a Spinoza prize grant [NWO SPI 92-266 to C.W.], a European Union Seventh Framework Programme grant (EU FP7) TANDEM project [HEALTH-F3-2012-305279 to C.W. and V.K.] and an NWO VENI

## Abstract

Candidaemia is a bloodstream infection caused by *Candida* species that primarily affects specific groups of at-risk patients. Because only small candidaemia patient cohorts are available, classical genome wide association cannot be used to identify *Candida* susceptibility genes. Therefore, we have applied an integrative genomics approach to identify novel susceptibility genes and pathways for candidaemia. *Candida*-induced transcriptome changes in human primary leukocytes were assessed by RNA sequencing. Genetic susceptibility to candidaemia was assessed using the Illumina immunochip platform for genotyping of a cohort of 217 patients. We then integrated genetics data with gene-expression profiles, *Candida*-induced cytokine production capacity, and circulating concentrations of cytokines. Based on the intersection of transcriptome pathways and genomic data, we prioritized 31 candidate genes for candidaemia susceptibility. This group of genes was enriched with genes involved in inflammation, innate immunity, complement, and hemostasis. We then validated the role of *MAP3K8* in cytokine regulation in response to *Candida* stimulation. Here, we present a new framework for the identification of susceptibility genes for infectious diseases that uses an unbiased, hypothesis-free, systems genetics approach. By applying this approach to candidaemia, we identified novel susceptibility genes and pathways for candidaemia, and future studies should assess their potential as therapeutic targets.

## Introduction

Genome-wide association studies (GWAS) have greatly contributed to the identification of susceptibility genes for human complex diseases. However, the need for large cohorts

grant [863.13.011 to Y.L.]. V.M. is supported by a PhD scholarship from Graduate School of Medical Sciences, University of Groningen, the Netherlands. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

precludes the use of GWAS to identify susceptibility genes for infectious diseases for which only relatively small patient cohorts can be recruited. So far, relatively few GWAS studies have been performed in patients with infections, including studies of viral and bacterial infections [1]. Compared to the hundreds of genetic loci that have been associated to human complex diseases, the number of susceptibility genes identified as associated to infectious diseases remains low. Given that the risk of death due to infectious diseases has a large genetic heritability [2], methods other than GWAS that can be applied to smaller cohorts are crucial to make progress in understanding and treating these diseases.

Candidaemia is the fourth most common systemic bloodstream infection in the United States (US) and is associated with mortality rates of up to 40% [3,4]. It is caused by opportunistic fungal pathogens belonging to the *Candida* species, particularly *Candida albicans* (*C. albicans)*, and primarily affects patients with a compromised immune system [5]. However, not all at-risk patients develop candidaemia, indicating that individual differences—including genetic background—influence their susceptibility to the infection. Despite the availability of potent antifungal drugs, the mortality rate of systemic *Candida* infections remains unacceptably high [6]. In addition to current treatment strategies, only adjuvant immunotherapy such as the administration of recombinant cytokines is believed to improve outcome [7]. Therefore, an understanding of the molecular pathways involved in the human host defence and identification of susceptibility genes is crucial for the design of appropriate prophylactic and immunotherapeutic strategies.

To identify genes underlying susceptibility to *Candida* infection, we have applied a systems genomics approach that integrates genetic data from candidaemia patients genotyped with the Immunochip platform [8] with *Candida*-induced gene-expression profiles in human leukocytes, cytokine production from *Candida*-stimulated peripheral blood mononuclear cells (PBMCs), and circulating cytokine concentrations in candidaemia patients. Using this approach, we demonstrate that genetic susceptibility loci with suggestive associations ($P < 9.99 \times 10^{-5}$) could play a major role in candidaemia susceptibility and we identify genes involved in inflammation, innate immunity, complement and hemostasis as having an important role in determining susceptibility to candidaemia.

## Materials and methods

### Study populations

To identify genetic variants associated with candidaemia, we performed a two-stage Immunochip-wide analysis of a candidaemia cohort using two control groups in 2014: a population-based healthy cohort and disease-matched cohort (European ancestry), as previously described [9]. Briefly, for Immunochip-wide association analysis, we first used a cohort consisting of 217 candidaemia patients of European ancestry and 11,920 population-based healthy controls (Discovery stage). The demographic and clinical characteristics of the candidaemia cohort have been previously described [9]. Re-analysis of the data and prioritization of genes from susceptibility loci with suggestive associations ($P < 9.99 \times 10^{-5}$) was performed in 2017.

After the discovery of single nucleotide polymorphisms (SNPs) for susceptibility to candidaemia using population-based healthy controls, we then validated our findings using a validation control consisting of 146 disease-matched but candidiasis-free controls. These candidiasis-free control patients were recruited from the same hospital wards as the candidaemia patients so that co-morbidities and clinical risk factors for infection were as similar as possible between patients and controls.

## Ethics statement

The study was approved by the institutional review boards at each study centre, and enrollment occurred between January 2003 and January 2009. The study centers are the Duke University Hospital (DUMC, Durham, North Carolina, USA) and Radboud University Nijmegen Medical Centre (Nijmegen, The Netherlands). All adult subjects provided informed written consent.

## Genotyping and case-control analysis of candidaemia cohort

DNA was isolated using the Gentra Pure Gene Blood kit (Qiagen, Venlo, the Netherlands) according to the protocol of the manufacturer. Genotyping of candidaemia patients and both control groups was performed using Immunochip according to Illumina's protocol [10]. Genotype data analysis and quality control of this cohort has been previously reported [9]. Briefly, in the discovery stage, the associations of Immunochip SNPs and susceptibility with candidaemia were tested by logistic regression using the first four components from the multidimensional scaling analysis as covariates. We considered a P value $< 9.99 \times 10^{-5}$ as the threshold for suggestive association to select 268 SNPs in 77 independent loci for validation. For validation analysis using candidaemia case-matched controls, the association between SNPs and candidaemia was tested by logistic regression using the first four multidimensional scaling analysis components as covariates. A validation P value $< 0.05$ was considered significant.

## PBMC isolation and stimulation with *Candida albicans*

Isolation of PBMCs from eight healthy volunteers and stimulation of PBMCs was performed as described previously [11]. After cell counting with a hemocytometer, the cell number was adjusted to $5 \times 10^6$/mL. To identify the transcriptome upon *Candida* stimulation, $5 \times 10^5$ isolated PBMCs were incubated with $1 \times 10^6$/mL heat-killed *C. albicans* (UC 820) (*C. albicans*: PBMC ratio of 1:2.5) and RPMI culture medium as a control for 4 and 24 hours. *C. albicans* UC 820 is a well-described strain regarding its immune responses in PBMCs [12].

## Analysis of RNA sequencing reads

The RNA sequencing analysis of this dataset was described previously [13]. Briefly, sequencing reads were mapped to the human genome using STAR (version 2.3.0). The aligner was provided with a file containing junctions from Ensembl GRCh37.71. The Htseq-count of the Python package HTSeq (version 0.5.4p3) was used (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html) to quantify the read counts per gene based on annotation version GRCh37.71, using the default union-counting mode. Differentially expressed genes were identified by statistical analysis using the DESeq2 package. The statistically significant threshold (False Discovery Rate $P < 0.05$ and Fold Change $\geq 2$) was applied.

## Pathway enrichment analysis

We performed gene enrichment analysis using ConsensusPathDB-human database (CPDB; http://cpdb.molgen.mpg.de/) [14]. The over-representation analysis is done using the default setting in which the database compares the predefined lists of functionally associated genes (pathways and Gene Ontology categories) to the list of differentially expressed genes and generates P values based on the hypergeometric test. The hypergeometric test P values are further corrected for multiple testing using the false discovery rate method.

### MAP3K8 inhibition of cytokines in *Candida*-stimulated PBMCs

PBMCs (5 x $10^5$) were placed in 96-well round-bottom plates in a final volume of 200μL. PBMCs were stimulated with (1 x $10^6$/mL) heat-inactivated conidia of *C. albicans* in the presence or absence of variable concentrations (10μM, 50μM, and 200μM) of a MAP3K8 inhibitor (#5240, Tocris Cookson Ltd., Bristol, UK) or a vehicle (DMSO) control. PBMCs were stimulated for 24 or 48 hours at 37˚C and 5% $CO_2$. After stimulation, culture supernatants were collected and stored at -20˚C until cytokine assays were performed. Cytokine levels from *Candida*-stimulated PBMCs were measured in cell culture supernatants using enzyme-linked immunosorbent assay (ELISA) according to the manufacturer's instructions (R&D Systems, MN, USA). Differences in cytokine levels were compared for statistical significance using the Wilcoxon rank test.

## Results

### *Candida albicans* induces transcription of genes involved in inflammation and immune-hemostasis interaction

In our experiments, we have used heat-killed *C. albicans* instead of live *C. albicans* for two main reasons. The first one is that long incubation periods of more than 48 hours to obtain the T cell derived cytokines IL-17, IL-22 and IFNγ result in overgrowth of the fungal cells and, ultimately, in cell lysis. Therefore, to achieve cell viability, we opted for heat-killed *C. albicans*. Secondly, the immune response is enhanced when using heat-killed *C. albicans* instead of live *C. albicans*. Heat-killing enhances the accessibility of β-glucans, whereas in live *C. albicans* yeast cells, β-glucans are shielded from immune recognition by a mannan layer. The β-glucans, which can be found in the inner layer of the *C. albicans* cell wall represent the main conserved pathogen-associated molecular patterns (PAMPs) that are recognized by one of the most important innate receptors for the recognition of *Candida* species, known as C-type lectin receptors (CLRs) [15]. The recognitions of β-glucans by CLRs constitutes the first step in the development of an immune response to *C. albicans*. Therefore, enhanced accessibility of β-glucans results in a pronounced immune response when using heat-killed *C. albicans* instead of live cells.

The transcriptome of PBMCs upon *Candida* stimulation was profiled by next-generation RNA sequencing (Fig 1). A total of 312 (4 hours) and 1,476 (24 hours) protein-coding genes were identified that showed >1.5-fold higher expression in cells stimulated with *C. albicans* compared to unstimulated cells (adjusted $P < 0.05$) (S1 Table, S1A and S1B Fig). A total of 246 protein-coding genes were found to be more strongly induced both 4 h and 24 h after stimulation (S2 Table), while 66 genes showed increased expression after 4 h only and 1,230 genes were more strongly expressed after 24 h only.

Among the significantly differentially expressed genes, pathway enrichment analysis showed that there is an overrepresentation of genes involved in cytokine signalling and of chemokine genes at both 4 h and 24 h after stimulation. Pathway enrichment analysis on differentially expressed protein-coding genes confirmed previously described *Candida*-response pathways including cytokine signalling (4 h: $P = 1.36 \times 10^{-12}$, 24 h: $P = 3.19 \times 10^{-9}$), Toll-like-receptor-mediated signalling [16] at 4 h (P = 0.00553), interferon signalling [17] (4 h: $P = 4.90 \times 10^{-10}$, 24 h: $P = 1.41 \times 10^{-5}$) and RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways (4 h: $P = 7.55 \times 10^{-6}$, 24 h: P = 0.00753) (S3 and S4 Tables) [18]. The enrichment analysis also uncovered an overrepresentation of differentially expressed genes involved in dissolution of the fibrin clot at 4 h (P = 0.0018) and 24 h (P = 0.00422), alternative complement activation at 4h (P = 0.0022), platelet activation, signalling and aggregation at 24 h (P = 0.0016) and activation of C3 and C5 -a central step of complement activation- at 24h (P = 0.00479). An even stronger effect for hemostasis ($P = 6.51 \times 10^{-6}$) was observed at 24h (Figs 2

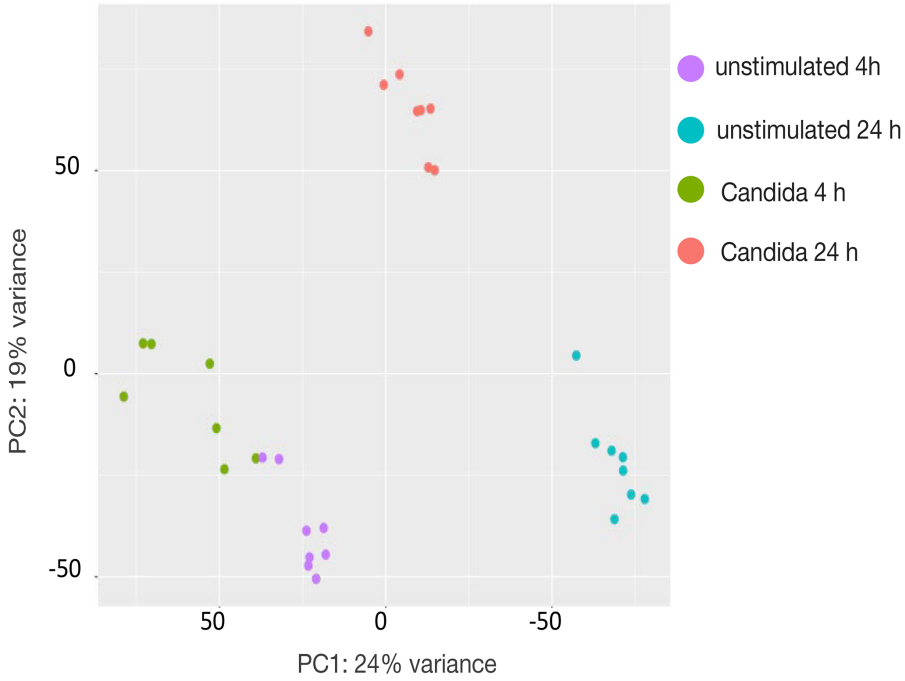



**Fig 1. Transcriptome of PBMCs upon *Candida* stimulation.** (A) Principal component analysis of RNA sequencing data obtained from eight healthy volunteers upon stimulation of their PBMCs with *C. albicans* for 4 or 24 hours or unstimulated. Four distinct groups were observed based on time-dependent exposure to *C. albicans* or culture medium alone using the first two principal components. Principal Component 1 (PC1, x-axis) represents 24% and PC2 (y-axis) represents 19% of total variation in the data. Green and red circles represent the stimulated samples with heat-killed *C. albicans* for 4 and 24 hours respectively and purple and blue circles represent the samples incubated with RPMI for 4 and 24 hours (used as controls) respectively.

https://doi.org/10.1371/journal.pone.0180824.g001

and 3). These results suggest that, along with genes involved in cytokine and interferon signalling, genes involved in the integration of immunity and hemostasis as biological processes may also be involved in the pathogenesis of systemic *Candida* infections.

## Immunochip-based genetic study identifies 18 candidaemia susceptibility loci

After identification of the transcriptome profile induced by *Candida* in human PBMCs, we explored whether these pathways could be validated through genetic association data from a clinical candidaemia cohort of 217 patients. We had previously reported three loci to be associated to *Candida* infection at genome-wide significance ($P < 5 \times 10^{-8}$) by performing a case-control study on the Immunochip platform [9]. In addition to these three loci, during the discovery stage, we identified 77 independent loci showing suggestive associations with P values lower than $9.99 \times 10^{-5}$. It is these independent loci that we have followed up in this study by

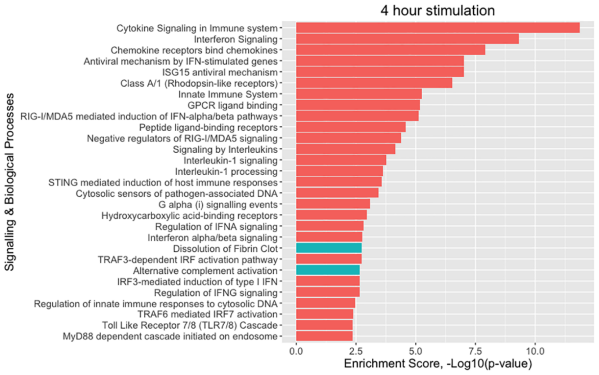



**Fig 2. Pathway enrichment analysis of the *Candida*-induced transcripts at 4 hours.** S3 Table displays all the signalling and biological processes as well as their enrichment p-values.

https://doi.org/10.1371/journal.pone.0180824.g002

validating them against 146 disease-matched controls to exclude confounding effects of the patient's clinical background. Following screening of the disease-matched controls, associations at 18 independent loci could be successfully validated ($P < 0.05$; S2 Fig and Table 1).

## eQTL and differential expression prioritize candidaemia causal genes enriched for inflammation and hemostasis

To identify causal genes from these 18 susceptibility loci, three different approaches were applied. In our first approach, we tested whether the 18 candidaemia SNPs were in linkage disequilibrium

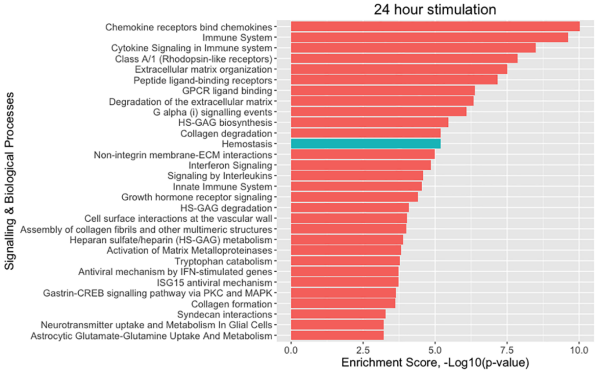



**Fig 3. Pathway enrichment analysis of the *Candida*-induced transcripts at 24 hours reveals a stronger effect for hemostasis at 24 hours than 4 hours.** S4 Table displays all the signalling and biological processes as well as their enrichment p-values.

https://doi.org/10.1371/journal.pone.0180824.g003

**Table 1. Prioritization of putative causal genes of candidaemia using an Immunochip-based association study.**

| | | | Discovery cohort | | Candidaemia cohort | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Chr** | **position** | **SNP** | **P value** | **OR** | **P value** | **OR** | **Risk allele** | **Frequency** | **Gene(s)** |
| 1 | 19199400 | rs6699706 | $1.01 \times 10^{-5}$ | 1.76 | $3.85 \times 10^{-2}$ | 1.58 | C | 0.11 | *ALDH4A1*[a] |
| 1 | 113793315 | rs11102637 | $5.86 \times 10^{-6}$ | 1.69 | $2.57 \times 10^{-3}$ | 1.84 | A | 0.15 | *MAGI3*[b] |
| 1 | 167833451 | rs3766122 | $3.01 \times 10^{-5}$ | 2.02 | $5.03 \times 10^{-3}$ | 3.22 | C | 0.04 | *F5*[a, c, d] |
| | | | | | | | | | *SELL*[a, d] |
| | | | | | | | | | *SELP*[d] |
| 1 | 199183753 | rs296537 | $1.57 \times 10^{-6}$ | 3.67 | $4.48 \times 10^{-2}$ | 4.74 | A | 0.01 | *IGFN1*[e] |
| | | | | | | | | | *LAD1*[e] |
| 2 | 127943178 | rs6748999 | $8.05 \times 10^{-6}$ | 2.21 | $6.23 \times 10^{-3}$ | 2.78 | G | 0.04 | *PROC*[c, d, e] |
| 3 | 50531585 | rs12491812 | $1.67 \times 10^{-5}$ | 3.24 | $2.48 \times 10^{-2}$ | 10.62 | T | 0.01 | *CISH*[e, f] |
| 5 | 33987450 | rs16891982 | $2.95 \times 10^{-5}$ | 2.14 | $1.48 \times 10^{-2}$ | 2.66 | C | 0.04 | *SLC45A2*[b] |
| 6 | 30078406 | rs11760176 | $9.76 \times 10^{-5}$ | 1.99 | $1.20 \times 10^{-3}$ | 3.74 | T | 0.05 | *ZNRD1*[a] |
| | | | | | | | | | *PPP1R11*[a] |
| 6 | 126875776 | rs1490387 | $9.80 \times 10^{-5}$ | 0.68 | $9.15 \times 10^{-3}$ | 0.66 | A | 0.46 | *CENPW*[b] |
| 9 | 116656377 | rs7022618 | $3.74 \times 10^{-5}$ | 1.64 | $3.74 \times 10^{-2}$ | 1.55 | C | 0.14 | *TNFSF15*[e] |
| | | | | | | | | | *TNFSF8*[a, e] |
| 9 | 122664970 | rs72758135 | $6.70 \times 10^{-5}$ | 0.45 | $5.77 \times 10^{-4}$ | 0.401 | C | 0.13 | *C5*[d] |
| | | | | | | | | | *PSMD5-AS1*[a] |
| | | | | | | | | | *STOM*[e] |
| 10 | 30763712 | rs1360119 | $3.64 \times 10^{-7}$ | 3.62 | $3.42 \times 10^{-2}$ | 5.13 | T | 0.01 | *MAP3K8*[e] |
| 10 | 54523962 | rs7092540 | $7.76 \times 10^{-5}$ | 2.50 | $3.01 \times 10^{-2}$ | 3.36 | A | 0.02 | *MBL2*[d] |
| 12 | 127841061 | rs59665078 | $7.03 \times 10^{-5}$ | 1.73 | $1.63 \times 10^{-2}$ | 1.85 | C | 0.09 | *GLT1D1*[e] |
| 14 | 94003143 | rs7149309 | $6.55 \times 10^{-5}$ | 2.84 | $3.34 \times 10^{-2}$ | 5.22 | T | 0.01 | *SERPINA1*[c, d, e] |
| | | | | | | | | | *IFI27*[e] |
| 16 | 28909510 | rs1802141 | $9.33 \times 10^{-7}$ | 3.06 | $2.26 \times 10^{-2}$ | 4.25 | G | 0.01 | *CD19*[d] |
| | | | | | | | | | *LAT*[c] |
| | | | | | | | | | *SPNS1*[f] |
| | | | | | | | | | *IL27*[e] |
| 17 | 74626807 | rs3848405 | $1.00 \times 10^{-6}$ | 3.32 | $1.28 \times 10^{-2}$ | 6.62 | C | 0.01 | *C1QTNF1*[e] |
| | | | | | | | | | *LGALS3BP*[e] |
| 19 | 50102284 | rs769450 | $6.63 \times 10^{-6}$ | 0.62 | $7.96 \times 10^{-4}$ | 0.57 | G | 0.4 | *TOMM40*[a] |
| | | | | | | | | | *BCL3*[e] |

Abbreviations: Chr, chromosome; OR, odds ratio.

[a] Candidaemia-associated SNPs showed an eQTL effect based on publicly available eQTL datasets.

[b] Genes in close proximity to candidaemia-associated SNP.

[c] These genes are involved in the blood coagulation pathway.

[d] These genes are involved in the complement system.

[e] Differentially expressed genes in response to *Candida* stimulation.

[f] Coding variants in the *CISH* and *SPNS1* genes were correlated with the candidaemia SNPs rs12491812 and rs1802141, respectively.

https://doi.org/10.1371/journal.pone.0180824.t001

(LD) with variants that alter the protein-coding sequence of the gene (both synonymous and non-synonymous). Two candidaemia-associated SNPs were identified, rs1802141 and rs12491812, that were in strong LD with synonymous variants (S5 Table) and therefore may point to putative causal genes. SNP rs1802141 on chromosome 16 is in strong LD ($R^2 = 0.82$, D' = 1) with a synonymous variant, rs61747536, in the *SPNS1* gene (Fig 4A). The *SPNS1* gene, also known as *HSpin1*, has been described to induce a caspase-independent autophagic cell death in cultured

human cells [19]. SNP rs12491812 on chromosome 3 is in strong LD ($R^2 = 0.89$, D' = 1) with two synonymous variants, rs2239753 and rs2239752, in the *CISH* gene (Fig 4B). The *CISH* locus has been associated with major infectious diseases such as bacteraemia, tuberculosis, malaria [20], and viral infections such as hepatitis B [21,22], suggesting shared susceptibility genes among different infections.

We should mention that the majority of candidaemia-associated SNPs and proxies fall within intergenic/intronic regions, suggesting that they could have a functional potential by affecting gene expression (S5 Table). Therefore, in our second approach, the 18 candidaemia-associated SNPs were mapped for *cis*-eQTLs using publicly available eQTL datasets from healthy blood donor samples [23–26]. *Cis*-eQTL mapping pinpointed 8 potential causal genes at 6 loci (Table 1).

In our third approach, genes that were located within a 500 kb window around the 18 candidaemia-associated SNPs were extracted and tested as to whether they are differentially expressed after *Candida* stimulation for 4 and 24 hours (S6 and S7 Tables). Within these 'susceptibility regions', we found 18 and 28 genes that are significantly induced upon *Candida*-stimulation for 4 and 24 hours respectively in human PBMCs, raising the possibility that some of these genes could be regulated by the candidaemia-associated SNPs in the context of *Candida* infection. For example, one of the 18 candidaemia-associated SNPs, rs7149309, is located within a locus on chromosome 14 that harbours a cluster of ten serine protease inhibitor genes of the serpin family (Fig 4C). Serpin Family A Member 1 (*SERPINA1*) was significantly expressed in PBMCs in response to *Candida* stimulation for 24 hours compared to the other *SERPIN* genes in the same locus (Fig 4D). *SERPINA1* encodes human alpha-1 antitrypsin (hAAT). A protective role of AAT has been demonstrated against different types of infectious pathogens, such as in bacterial peritonitis [27] and pulmonary *Pseudomonas aeruginosa* infection in cystic fibrosis patients [28].

Together, our three approaches prioritized 31 putative candidate genes for candidaemia susceptibility located in 18 loci (Table 1). Intriguingly, nine out of these 31 prioritized genes (*LAT*, CD19, *F5*, *PROC*, *C5*, *SERPINA1*, *SELP*, *SELL* and *SELE*) are enriched in the processes of complement and blood coagulation (Table 1, S3 Fig and S8 Table). As expected, pathway enrichment analysis for all 31 candidaemia genes showed enrichment for cytokine- and immune-related signalling pathways, which is in agreement with our pathway enrichment analysis of the *Candida*-induced transcripts at 4 and 24 hours (Figs 2 and 3, S3 and S4 Tables). In three out of 18 loci, genes that are located in close proximity to the top SNP were shown (Table 1), as we did not find any evidence from the above three approaches to prioritize causal genes.

## *MAP3K8* modulates cytokine production in patients and in experimental models

One of the candidaemia-associated SNPs, rs1360119, was mapped to the Mitogen-Activated Protein Kinase Kinase 8 (*MAP3K8*) locus (Fig 5A). In particular, rs1360119 is an intronic variant, suggesting that it may affect gene expression. However, rs1360119 does not show *cis*-eQTL effect using publicly available eQTL datasets from healthy blood donor samples (Table 1). That does not exclude the fact that this SNP may show an eQTL effect under context-specific conditions, for instance, upon *Candida* stimulation. In addition, an intronic variant can have functional effects on splicing and, therefore, we can speculate that this SNP may affect splicing.

Of note, MAP kinases are notably attractive therapeutic targets [29]. Therefore, considering the regulatory role of MAP3K8 protein in the production of cytokines in response to lipopolysaccharide [30], bacterial [31,32], and viral pathogens [33], we tested whether the rs1360119 SNP

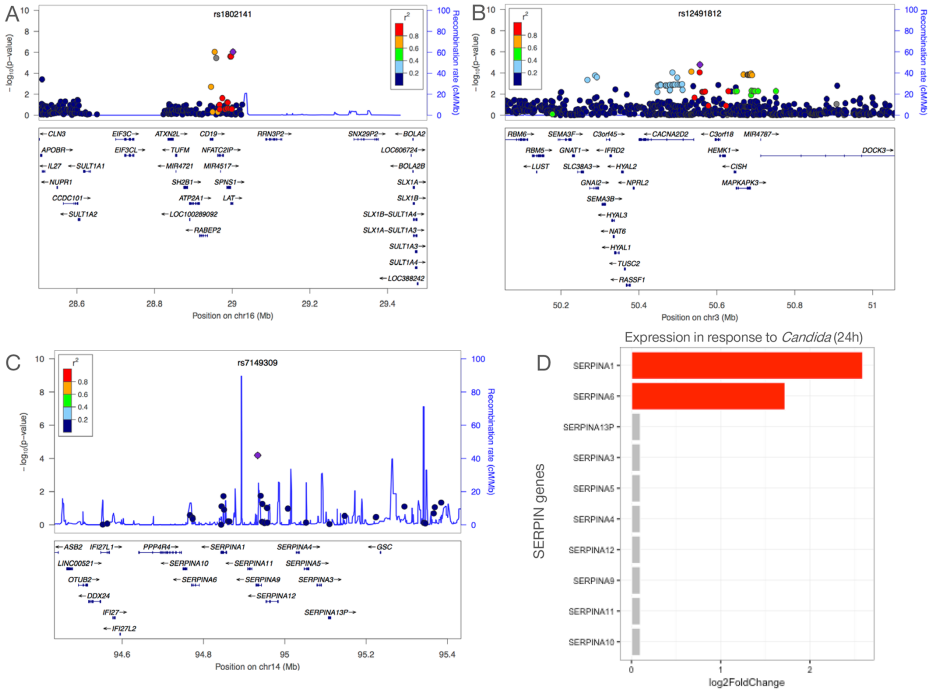



**Fig 4. Prioritization of candidaemia genes enriched for blood coagulation based on eQTL and differential expression data.** Regional association plots are shown around candidaemia-associated SNPs identified in the Discovery cohort. (A) rs1802141 ($P = 9.325 \times 10^{-7}$) on chromosome 16 is in strong LD with the synonymous variant rs61747536 ($R^2 = 0.82$, D' = 1) in the *SPNS1* gene and (B) rs12491812 ($P = 1.67 \times 10^{-5}$) on chromosome 3 is in strong LD ($R^2 = 0.89$, D' = 1) with two synonymous variants, rs2239753 and rs2239752, in the *CISH* gene. SNPs are plotted as the–log10 of the p-value. Local LD structure is reflected by the plotted estimated recombination rates (from HapMap) in the region around the associated SNP (purple diamond) and its correlated proxies. The correlation of the lead SNP to other SNPs at the locus is indicated by colour. (C) Regional association plot for rs7149309, which is located in a locus on chromosome 14, that harbors a cluster of serine protease inhibitor genes of the serpin family, with *SERPINA1* being the most likely causal gene at this locus based on our expression data upon *Candida* stimulation. (D) Log2 fold change expression levels of all *SERPIN* genes upon *Candida* stimulation for 24 hours at a locus on chromosome 14 marked by the candidaemia-associated SNP rs7149309. *SERPINA1* showed a log2 fold change of ~2.6.

https://doi.org/10.1371/journal.pone.0180824.g004

correlates with levels of pro-inflammatory cytokines in serum of candidaemia patients. The T allele of the SNP rs1360119 is associated with decreased IL-6 (P = 0.018), IL-8 (P = 0.024), and interferon (IFNγ) (P = 0.013) cytokine levels (Figs 5B–5D). Five other candidaemia-associated SNPs showed a moderate association ($P < 0.05$) with circulating cytokine levels (S9 Table), suggesting that some of the candidaemia-associated variants could determine susceptibility to candidaemia by influencing cytokine production capacity.

To test whether *MAP3K8* is involved in regulating *Candida*-induced cytokine production, we measured cytokine levels upon stimulation of PBMCs with *C. albicans* in the presence or absence of a MAP3K8 chemical inhibitor. Three different cytokines (TNFα, IL-6,

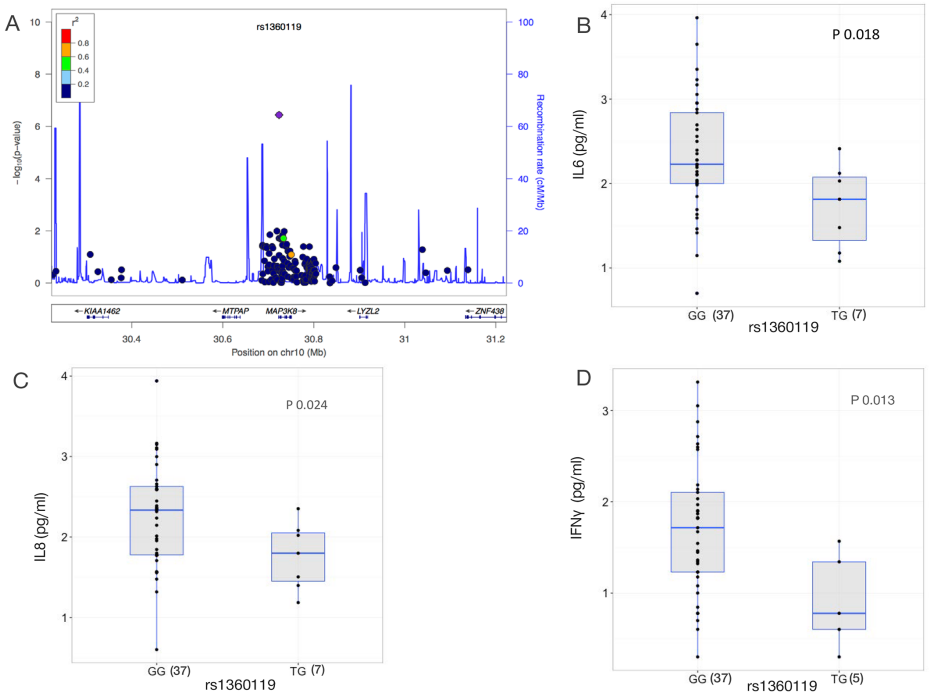



**Fig 5. Candidaemia-associated SNP rs1360119 affects the cytokine expression in serum from patients.** (A) Regional association plot for candidaemia-associated SNP rs1360119, which was mapped to *MAP3K8* locus on chromosome 10. (B) Genotypes of candidaemia-associated SNP rs1360119 are consistently associated with expression levels of (B) IL6 (P = 0.018), (C) IL8 (P = 0.024) and (D) IFNɣ (P = 0.013) as measured in serum of candidaemia patients. The numbers of individuals per genotype are shown in parentheses. Cytokine levels were log transformed and P values were obtained using Kruskal Wallis test, and 0.05 was considered statistically significant (minor allele frequency of A = 2% in Europeans from 1000 Genomes Project Phase 3, http://www.internationalgenome.org).

https://doi.org/10.1371/journal.pone.0180824.g005

and IFNγ) known to be involved in host defence against *C. albicans* were measured [15]. MAP3K8 activity was blocked using three different concentrations of the MAP3K8 inhibitor (10 μM, 50 μM, and 200 μM). Overall, we observed a dose-dependent reduction in all three cytokine levels upon increasing concentration of MAP3K8 inhibitor. A significant decrease in *Candida*-induced TNFα production was observed compared to the DMSO-control at 200 μM of *MAP3K8* inhibitor (P = 0.03) upon 24-hour stimulation (Fig 6A). In addition, a significant decrease in IL-6 was observed upon stimulation with *Candida* conidia for 24 hours with 200 μM *MAP3K8* inhibitor compared to DMSO (P = 0.04) (Fig 6B). We should note that IFNγ, one of the most crucial cytokines for efficient host defence for systemic candidiasis, decreased compared to DMSO after stimulation with *Candida* conidia (P = 0.057) for 48 hours at 200 μM of *MAP3K8* inhibitor (Fig 6C). Overall, these experiments demonstrate that *MAP3K8* regulates cytokine levels upon *Candida* infection.

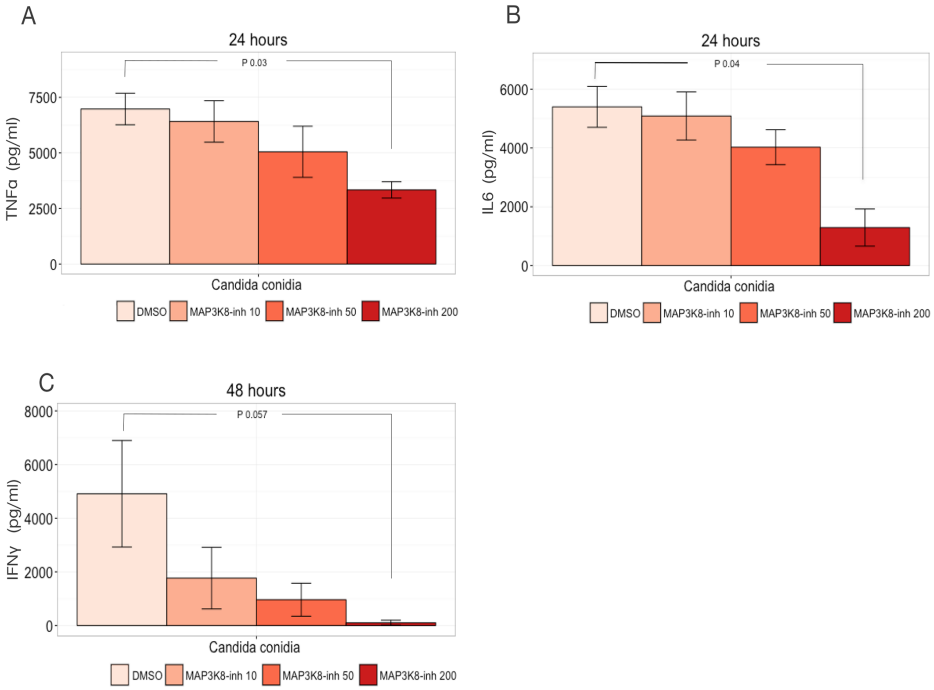



**Fig 6. *MAP3K8* modulates cytokine production in PBMCs.** Median expression levels of (A) TNFα (B) IL6 and (C) IFNγ in *Candida*-stimulated PBMCs upon inhibition of MAP3K8 at three different concentrations (10 μM, 50 μM, and 200 μM). *Candida conidia* was added at a concentration of 1 x $10^6$/ml. P values were obtained using Wilcoxon rank test (*P < 0.05). Data shown are from two independent experiments of PBMC stimulation for 24 and 48 hours with conidia of *C. albicans*.

https://doi.org/10.1371/journal.pone.0180824.g006

## Discussion

Although *Candida spp* are the fourth most common cause of sepsis in the US, and the most common cause of fungal sepsis in Europe, the difficulty of recruiting sufficient numbers of patients has prevented the use of large GWAS studies to reveal the genetic factors involved in the pathophysiology of *Candida* infections. Our study demonstrates the potential of using a systems genomics approach to obtain novel insights into the genetic basis and host defence mechanisms involved in relatively rare infections in which classical GWAS studies are not applicable. By integrating transcriptomic, genetic and immunological studies, we have identified novel susceptibility genes for candidaemia and new potential therapeutic targets.

The first major group of pathways that are induced during stimulation of human PBMCs with *C. albicans* are genes important for inflammation and innate immunity, cytokine and chemokine synthesis, interferon and inflammasome signalling. This is not surprising as these pathways have been previously shown to have a crucial role for antifungal host defence [15]. The importance of inflammation and innate immunity was also validated by genetic studies

that confirmed that the genes determining susceptibility to candidaemia were enriched for inflammatory genes. However, our analysis has also provided new insight in additional processes that apparently have an important impact on candidaemia. One of these processes is complement activation, which has been shown to modulate cytokine production induced by *Candida* stimulation [34].

Hemostasis is another important biological process suggested by our study to be involved in susceptibility to candidaemia. Hemostasis is known to be strongly activated during sepsis and to interact with the inflammatory responses [35–38]. Strong interactions between immune defence and hemostasis are well documented for bacterial infections, and hemostasis has been previously associated with an increased susceptibility for bacterial sepsis [37,39,40]. Some of the hemostasis genes that we identified as candidaemia susceptibility genes (*SERPINA1*, *LAT* and *F5*) have also been shown to be important for bacterial sepsis (S4A Fig) [41]. Future studies should examine the involvement of these pathways and mechanisms in systemic *Candida* infection, especially as platelets contain several anti-fungal defensins [42], indicating their potential involvement in candidaemia. For example, an examination of protein-protein interaction data using the Plateletweb database found that 14 of our candidaemia susceptibility proteins were found in platelets (S4B Fig).

Another novel observation of the present study is the association of *SERPINA1* gene with susceptibility to candidaemia. Mutations in this gene lead to AAT deficiency and predispose individuals to chronic obstructive pulmonary disease and liver diseases [43,44]. Considering the important role of AAT in modulation of inflammation, as well as the recent beneficial effect of human AAT against bacterial infections [45], hAAT could represent a potential novel adjuvant immunotherapy against systemic *Candida* infection.

*MAP3K8* is known as a critical gene in innate immune responses linking pattern recognition receptors such as Toll-like receptors (TLRs) to TNF production through activation of extracellular signal-regulated kinase [30,46]. In this study we also demonstrate that *MAP3K8* plays an important role in *Candida* infection by modulating cytokine production. The link of *MAP3K8* with TLR-signalling suggests that the innate immune response is important in the context of *Candida* infection. In particular, non-synonymous SNPs in TLR1 have previously been associated with increased susceptibility to candidaemia, highlighting the role of TLRs in the recognition of *Candida* species [47]. Furthermore, the critical role of *MAP3K8* in host immune defence has been demonstrated for other infectious pathogens including influenza virus [33] and *Listeria monocytogenes* [32]. Most importantly, MAP3K8 is an important and novel therapeutic target for inflammatory diseases [48]. A recent computational approach using publicly available transcriptome datasets for the discovery of common immunomodulators in fungal infections also pinpointed *MAP3K8* and *SERPINE1* in the top ten consistently perturbed gene sets [49]. However, further functional studies are needed to shed light on the role of these genes in host defence against *Candida* infection and to understand their potential as therapeutic targets.

Several limitations also apply to this study. First, the patient cohort study is relatively small, which limited our power to identify many of the genetic factors influencing susceptibility to candidaemia. Second, we lack genetic validation in an independent cohort of patients. This is because no such cohort is currently available: the cohort studied here is the largest candidaemia cohort currently available. In addition, a potential limitation of the present study is that the Immunochip platform covers only 5% of the human genome, i.e. we still lack genome-wide information. This means that further critical genetic variations for candidaemia remain to be discovered on a genome-wide scale. Furthermore, it should be mentioned that *C.albicans* is a polymorphic fungus and is encountered either as yeast or hyphal forms. The transition between conidia and hyphae is a virulence trait of *C. albicans* and mutants that are locked in

the yeast form are less virulent in experimental models of disseminated candidiasis [50,51]. Differential cytokine expression and excretion by immune cells may explain the increased invasiveness of hyphae. For instance, hyphae form has been shown to be unable to induce IFNγ in either human PBMCs or murine splenic lymphocytes and conidia induced a much higher TNFα production than hyphae did [52]. This differential cytokine expression may be attributed to structural differences in the cell wall between yeast and hyphal forms and therefore, the human innate immune system can discriminate between yeast and hyphae [53]. In addition, the accessibility of different pathogen associated molecular patterns (PAMPs) between hyphae and conidia of *C. albicans* as well as the potential of only hyphae to activate the inflammasome can explain the induction of immune responses that discriminate between conidia and hyphae forms [53,54]. Thus, by using only heat-killed *C. albicans* conidia in the present study may not represent the full physiological conditions in humans, and further studies should address the differential effect of both forms in PBMCs. Last, *C. albicans* strains may vary in pathogenicity and, therefore, may elicit different host immune responses making it interesting to consider more clinical strains (in addition to UC 820) for validation in the future.

To conclude, the application of an unbiased, hypothesis-free, systems integrative genomics approach has the power to identify novel susceptibility genes for infectious diseases such as systemic candidiasis. In this study, this approach highlighted genes in inflammation, innate immunity and hemostasis pathways that contribute to the genetic susceptibility against candidaemia. We therefore believe that such integrative approaches are an important tool for future identification of genetic susceptibility to rare infectious diseases.

## Supporting information

**S1 Fig.** Heatmaps showing the expression of protein-coding genes, which showed >1.5-fold higher expression, upon (A) 4 and (B) 24-hour stimulation with *C. albicans* in PBMCs from healthy volunteers. RPMI medium was used as control. (adjusted $P < 0.05$).
(TIF)

**S2 Fig. Immunochip-wide association analysis with candidaemia.** Manhattan plot highlighting the 18 independent loci showing suggestive association with candidaemia ($P < 9.99 \times 10^{-5}$) using a second set of case-matched controls. The y-axis represents the–$\log_{10}$P values of 122,779 SNPs. Their chromosomal positions are shown on the x axis. The dotted line represents the suggestive threshold for association ($P < 9.99 \times 10^{-5}$). P values were not corrected for multiple testing when testing for association with candidaemia susceptibility at 18 independent loci identified in the discovery stage.
(TIF)

**S3 Fig. Pathway enrichment analysis based on KEGG and Reactome sources of all 31 candidaemia genes prioritized based on eQTL, differential expression upon *Candida* stimulation and proximity to top SNP.** Candidaemia genes showed an expected enrichment for cytokine signalling pathways and showed a strong enrichment for complement and coagulation pathways. Each node represents a separate pathway whose number of genes and P-value are encoded as node size and node colour, respectively. Two nodes are connected by an edge if they share members. The edge width reflects the relative overlap (corresponding to the Fowlkes-Mallows index) between the nodes, while the edge colour encodes the number of shared gene members. (see.tif image)
(TIF)

**S4 Fig.** (A) Heatmap depicts the false discovery rate (FDR) of differentially expressed genes in sepsis as identified by *Davenport E et al* in their discovery and validation cohort. These genes

were differentially expressed in response to *Candida* stimulation as well. (B) Proteins encoded by 14 candidaemia susceptibility genes detected in platelets using plateletWeb (http://plateletweb.bioapps.biozentrum.uni-wuerzburg.de/plateletweb.php). (see.tiff image)
(TIF)

**S1 Table. Differentially expressed protein-coding genes in response to 4 and 24 hour-*Candida* stimulation that showed >1.5-fold higher expression compared to RPMI medium used as control.**
(DOCX)

**S2 Table. A total of 246 protein-coding genes were differentially expressed at both 4 and 24 hour-*Candida* stimulation, showing >1.5-fold higher expression compared to RPMI medium used as control.**
(DOCX)

**S3 Table. Pathway enrichment analysis on differentially expressed protein-coding genes that showed >1.5 higher expression compared to RPMI medium in response to 4 hour-*Candida* stimulation.**
(DOCX)

**S4 Table. Pathway enrichment analysis on differentially expressed protein-coding genes that showed >1.5 higher expression compared to RPMI medium in response to 24 hour-*Candida* stimulation.**
(DOCX)

**S5 Table. Candidaemia-associated SNPs and variants with $r^2 > = 0.8$.** SNPs rs12491812 and rs1802141 were in strong linkage disequilibrium (LD) with synonymous variants located in *CISH* and *SNPS1* genes respectively. (source: Haploreg http://archive.broadinstitute.org/mammals/haploreg/haploreg.php).
(DOCX)

**S6 Table. Differential expression of genes that are located within a 500 kilobase (kb) window around the candidaemia-associated SNPs upon *Candida* stimulation at 4 hours.** Bolded genes show a log2 fold change > 1.5.
(DOCX)

**S7 Table. Differential expression of genes that are located within a 500 kilobase (kb) window around the candidaemia-associated SNPs upon *Candida* stimulation at 24 hours.** Bolded genes show a log2 fold change > 1.5.
(DOCX)

**S8 Table. All thirty-one prioritized susceptibility genes for candidaemia showed a strong enrichment for complement and coagulation pathways along with cytokine- and immune-related pathways based on KEGG and Reactome sources.**
(DOCX)

**S9 Table. Five additional candidaemia SNPs showed a moderate association with circulating cytokine levels as measured in serum from candidaemia patients.** Cytokine levels were log transformed and statistical significance was tested with Kruskal Wallis test.
(DOCX)

## Acknowledgments

The authors would like to thank Kate Mc Intyre for editing the manuscript.

## Author Contributions

**Conceptualization:** Cisca Wijmenga, Mihai G. Netea, Vinod Kumar.

**Data curation:** Vasiliki Matzaraki, Vinod Kumar.

**Formal analysis:** Vasiliki Matzaraki, Isis Ricaño-Ponce, Lude Franke, Sebo Withoff, Yang Li, Vinod Kumar.

**Funding acquisition:** Cisca Wijmenga, Mihai G. Netea, Vinod Kumar.

**Investigation:** Vasiliki Matzaraki, Mark S. Gresnigt, Martin Jaeger, Melissa D. Johnson, Marije Oosting, John R. Perfect, Leo A. B. Joosten, Bart Jan Kullberg, Frank L. van de Veerdonk, Mihai G. Netea, Vinod Kumar.

**Methodology:** Vasiliki Matzaraki, Yang Li, Vinod Kumar.

**Project administration:** Cisca Wijmenga, Mihai G. Netea, Vinod Kumar.

**Resources:** Melissa D. Johnson, John R. Perfect, Leo A. B. Joosten, Bart Jan Kullberg, Frank L. van de Veerdonk, Cisca Wijmenga, Mihai G. Netea, Vinod Kumar.

**Supervision:** Cisca Wijmenga, Mihai G. Netea, Vinod Kumar.

**Validation:** Vasiliki Matzaraki, Melissa D. Johnson, John R. Perfect, Leo A. B. Joosten, Bart Jan Kullberg, Frank L. van de Veerdonk.

**Visualization:** Vasiliki Matzaraki, Yang Li, Vinod Kumar.

**Writing – original draft:** Vasiliki Matzaraki, Cisca Wijmenga, Mihai G. Netea, Vinod Kumar.

**Writing – review & editing:** Mark S. Gresnigt, Martin Jaeger, Melissa D. Johnson, Sebo Withoff, John R. Perfect, Bart Jan Kullberg, Iris Jonkers, Mihai G. Netea, Vinod Kumar.

## References


1. Newport MJ, Finan C. Genome-wide association studies and susceptibility to infectious diseases. Brief. Funct. Genomics. 2011; 10:98–107. https://doi.org/10.1093/bfgp/elq037 PMID: 21436306
2. Obel N, Christensen K, Petersen I, Sørensen TIA, Skytthe A. Genetic and environmental influences on risk of death due to infections assessed in Danish twins, 1943–2001. Am. J. Epidemiol. 2010; 171:1007–13. https://doi.org/10.1093/aje/kwq037 PMID: 20375195
3. Wisplinghoff H, Bischoff T, Tallent SM, Seifert H, Wenzel RP, Edmond MB. Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. Clin. Infect. Dis. 2004; 39:309–17. https://doi.org/10.1086/421946 PMID: 15306996
4. Campion EW, Kullberg BJ, Arendrup MC. Invasive Candidiasis. N. Engl. J. Med. 2015; 373:1445–56. https://doi.org/10.1056/NEJMra1315399 PMID: 26444731
5. Wisplinghoff H, Seifert H, Wenzel RP, Edmond MB. Inflammatory response and clinical course of adult patients with nosocomial bloodstream infections caused by Candida spp. Clin. Microbiol. Infect. 2006; 12:170–7. https://doi.org/10.1111/j.1469-0691.2005.01318.x PMID: 16441456
6. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden Killers: Human Fungal Infections. Sci. Transl. Med. 2012; 4:1–9.
7. van de Veerdonk FL, Kullberg BJ, Netea MG. Adjunctive immunotherapy with recombinant cytokines for the treatment of disseminated candidiasis. Clin. Microbiol. Infect. 2012; 18:112–9. https://doi.org/10.1111/j.1469-0691.2011.03676.x PMID: 22032929
8. Cortes A, Brown M a. Promise and pitfalls of the Immunochip. Arthritis Res. Ther. 2011; 13:101. https://doi.org/10.1186/ar3204 PMID: 21345260
9. Kumar V, Cheng SC, Johnson MD, Smeekens SP, Wojtowicz A, Giamarellos-Bourboulis E, et al. Immunochip SNP array identifies novel genetic variants conferring susceptibility to candidaemia. Nat. Commun. 2014; 5:4675. https://doi.org/10.1038/ncomms5675 PMID: 25197941

10. Trynka G, Hunt K a, Bockett N a, Romanos J, Castillejo G, Concha EG De, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. 2012; 43:1193–201.
11. Netea MG, Gow NAR, Munro CA, Bates S, Collins C, Ferwerda G, et al. Immune sensing of *Candida albicans* requires cooperative recognition of mannans and glucans by lectin and Toll-like receptors. J. Clin. Invest. 2006; 116:1642–50. https://doi.org/10.1172/JCI27114 PMID: 16710478
12. Lehrer RI, Cline MJ. Interaction of *Candida albicans* with human leukocytes and serum. J. Bacteriol. 1969; 98:996–1004. PMID: 4182532
13. Li Y, Oosting M, Deepen P, Ricaño-Ponce I, Smeekens S, Jaeger M, et al. Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. Nat. Med. 2016; 22:952–60. https://doi.org/10.1038/nm.4139 PMID: 27376574
14. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 Update. Nucleic Acids Res. 2013; 41 (Database issue):D793–D800. https://doi.org/10.1093/nar/gks1055 PMID: 23143270
15. Netea MG, Joosten LA, van der Meer JW, Kullberg BJ, van de Veerdonk FL. Immune defence against *Candida* fungal infections. Nat Rev Immunol. 2015; 15:630–42. https://doi.org/10.1038/nri3897 PMID: 26388329
16. Gow N a. R, van de Veerdonk FL, Brown AJP, Netea MG. *Candida albicans* morphogenesis and host defence: discriminating invasion from colonization. Nat. Rev. Microbiol. 2011; 10:112–22. https://doi.org/10.1038/nrmicro2711 PMID: 22158429
17. Smeekens SP, Ng A, Kumar V, Johnson MD, Plantinga TS, van Diemen C, et al. Functional genomics identifies type I interferon pathway as central for host defense against *Candida albicans*. Nat Commun. 2013; 4:1342. https://doi.org/10.1038/ncomms2343 PMID: 23299892
18. Jaeger M, van der Lee R, Cheng SC, Johnson MD, Kumar V, Ng A, et al. The RIG-I-like helicase receptor MDA5 (IFIH1) is involved in the host defense against *Candida* infections. Eur. J. Clin. Microbiol. Infect. Dis. 2015; 34:963–74. https://doi.org/10.1007/s10096-014-2309-2 PMID: 25579795
19. Yanagisawa H, Miyashita T, Nakano Y, Yamamoto D. HSpin1, a transmembrane protein interacting with Bcl-2/Bcl-xL, induces a caspase-independent autophagic cell death. Cell Death Differ. 2003; 10:798–807. https://doi.org/10.1038/sj.cdd.4401246 PMID: 12815463
20. Wang Y, Wang W. CISH and susceptibility to infectious diseases. N. Engl. J. Med. 2010; 363:1676; author reply 1676–7.
21. Hu Z, Yang J, Wu Y, Xiong G, Wang Y, Yang J, et al. Polymorphisms in *CISH* gene are associated with persistent hepatitis B virus infection in Han Chinese population. PLoS One. 2014; 9(6): e100826. https://doi.org/10.1371/journal.pone.0100826 PMID: 24964072
22. Tong H V., Toan NL, Song LH, Kremsner PG, Kun JFJ, Velavan TP. Association of CISH-292A/T genetic variant with hepatitis B virus infection. Immunogenetics. 2012; 64:261–5. https://doi.org/10.1007/s00251-011-0584-y PMID: 22033525
23. Ward LD, Kellis M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40 (Database issue):D930–4. https://doi.org/10.1093/nar/gkr917 PMID: 22064851
24. Bonder MJ, Luijk R, Zhernakova D V, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017; 49:131–138. https://doi.org/10.1038/ng.3721 PMID: 27918535
25. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 2013; 45:1238–43. https://doi.org/10.1038/ng.2756 PMID: 24013639
26. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348:648–60. https://doi.org/10.1126/science.1262110 PMID: 25954001
27. Kaner Z, Ochayon DE, Shahaf G, Baranovski BM, Bahar N, Mizrahi M, et al. Acute phase protein α1-antitrypsin reduces the bacterial burden in mice by selective modulation of innate cell responses. J. Infect. Dis. 2015; 211:1489–98. https://doi.org/10.1093/infdis/jiu620 PMID: 25389308
28. Griese M, Latzin P, Kappler M, Weckerle K, Heinzlmaier T, Bernhardt T, et al. alpha1-Antitrypsin inhalation reduces airway inflammation in cystic fibrosis patients. Eur. Respir. J. 2007; 29:240–50. https://doi.org/10.1183/09031936.00047306 PMID: 17050563
29. Patterson H, Nibbs R, Mcinnes I, Siebert S. Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases. Clin. Exp. Immunol. 2014; 176:1–10. https://doi.org/10.1111/cei.12248 PMID: 24313320

30. Dumitru CD, Ceci JD, Tsatsanis C, Kontoyiannis D, Stamatakis K, Lin JH, et al. TNF-alpha induction by LPS is regulated posttranscriptionally via a Tpl2/ERK-dependent pathway. Cell. 2000; 103:1071–83. PMID: 11163183

31. Xiao N, Eidenschenk C, Krebs P, Brandl K, Blasius AL, Xia Y, et al. The Tpl2 mutation Sluggish impairs type I IFN production and increases susceptibility to group B streptococcal disease. J. Immunol. 2009; 183:7975–83. https://doi.org/10.4049/jimmunol.0902718 PMID: 19923465

32. Mielke L a Elkins KL, Wei L, Starr R, Tsichlis PN, O'Shea JJ, et al. Tumor progression locus 2 (Map3k8) is critical for host defense against Listeria monocytogenes and IL-1 beta production. J. Immunol. 2009; 183:7984–93. https://doi.org/10.4049/jimmunol.0901336 PMID: 19933865

33. Kuriakose T, Tripp RA, Watford WT. Tumor progression locus 2 promotes induction of IFNλ, interferon stimulated genes and antigen-specific $CD8^+$ T cell responses and protects against influenza virus. PLoS Pathog. 2015; 11:1–22.

34. Zipfel PF, Skerka C. Complement, *Candida*, and cytokines: The role of C5a in host response to fungi. Eur. J. Immunol. 2012; 42(4):822–5. https://doi.org/10.1002/eji.201242466 PMID: 22531909

35. Delvaeye M, Conway EM. Coagulation and innate immune responses: Can we view them separately? Blood. 2009; 114(12):2367–74. https://doi.org/10.1182/blood-2009-05-199208 PMID: 19584396

36. Esmon CT, Xu J, Lupu F. Innate immunity and coagulation. J. Thromb. Haemost. 2011; 9:182–8. https://doi.org/10.1111/j.1538-7836.2011.04323.x PMID: 21781254

37. Esmon CT. The interactions between inflammation and coagulation. Br. J. Haematol. 2005; 131 (4):417–30. https://doi.org/10.1111/j.1365-2141.2005.05753.x PMID: 16281932

38. Aird WC. Sepsis and coagulation. Crit. Care Clin. 2005; 21(3):417–31. https://doi.org/10.1016/j.ccc.2005.04.004 PMID: 15992665

39. Khakpour S, Wilhelmsen K, Hellman J. Vascular endothelial cell Toll-like receptor pathways in sepsis. Innate Immun. 2015; 21:827–46. https://doi.org/10.1177/1753425915606525 PMID: 26403174

40. Opal SM EC. Bench-to-bedside review: functional relationships between coagulation and the innate immune response and their respective roles in the pathogenesis of sepsis. Crit Care. 2003; 7(1):23–38. https://doi.org/10.1186/cc1854 PMID: 12617738

41. Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, et al. Genomic landscape of the individual host response and outcomes in sepsis: A prospective cohort study. Lancet Respir. Med. 2016; 4:259–71. https://doi.org/10.1016/S2213-2600(16)00046-1 PMID: 26917434

42. Speth C, Rambach G, Lass-Flörl C. Platelet immunology in fungal infections. Thromb. Haemost. 2014; 112:632–9. https://doi.org/10.1160/TH14-01-0074 PMID: 24990293

43. Stoller JK, Aboussouan LS. A review of a1-antitrypsin deficiency. Am. J. Respir. Crit. Care Med. 2012; 185:246–59. https://doi.org/10.1164/rccm.201108-1428CI PMID: 21960536

44. Lomas DA, Mahadeva R. α1-antitrypsin polymerization and the serpinopathies: Pathobiology and prospects for therapy. J. Clin. Invest. 2002; 110:1585–90. https://doi.org/10.1172/JCI16782 PMID: 12464660

45. Kanera Ziv, Ochayona David E., Shahaf Galit, Baranovski Boris M., Bahar Nofar, Mark Mizrahi ECL. Acute Phase Protein α1-Antitrypsin Reduces the Bacterial Burden in Mice by Selective Modulation of Innate Cell Responses. J Infect Dis. 2015; 211:1489–98. https://doi.org/10.1093/infdis/jiu620 PMID: 25389308

46. Waterfield MR, Zhang M, Norman LP, Sun SC. NF-kB1/p105 regulates lipopolysaccharide-stimulated MAP kinase signaling by governing the stability and function of the Tpl2 kinase. Mol. Cell. 2003; 11:685–94. PMID: 12667451

47. Plantinga TS, Johnson MD, Scott WK, Van De Vosse E, Velez Edwards DR, Smith PB, et al. Toll-like receptor 1 polymorphisms increase susceptibility to candidemia. J. Infect. Dis. 2012; 205:934–43. https://doi.org/10.1093/infdis/jir867 PMID: 22301633

48. George D, Salmeron A. Cot/Tpl-2 protein kinase as a target for the treatment of inflammatory disease. Curr Top Med Chem. 2009; 9(7):611–622. PMID: 19689369

49. Kidane YH, Lawrence C, Murali TM. Computational approaches for discovery of common immunomodulators in fungal infections: towards broad-spectrum immunotherapeutic interventions. BMC Microbiol. 2013; 13:224. https://doi.org/10.1186/1471-2180-13-224 PMID: 24099000

50. Krueger KE, Ghosh AK, Krom BP, Cihlar RL. Deletion of the NOT4 gene impairs hyphal development and pathogenicity in *Candida albicans*. Microbiology. 2004; 150:229–40. https://doi.org/10.1099/mic.0.26792-0 PMID: 14702416

51. Warenda AJ, Kauffman S, Sherrill TP, Becker JM, Konopka JB. *Candida albicans* septin mutants are defective for invasive growth and virulence. Infect. Immun. 2003; 71:4045–51. https://doi.org/10.1128/IAI.71.7.4045-4051.2003 PMID: 12819094

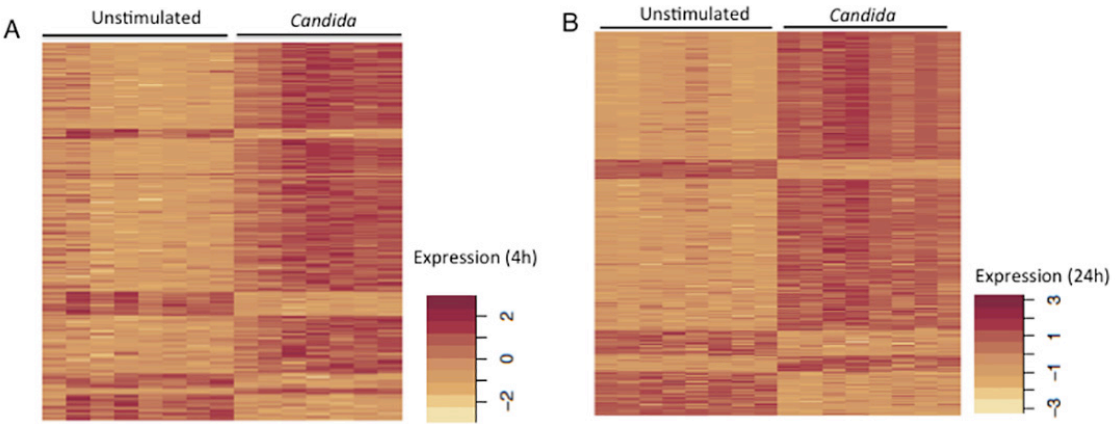
52. Graaf C a a Van Der, Netea MG, Meer, Kullberg BJ, Verschueren I. Differential cytokine production and Toll-like receptor signaling pathways by *Candida albicans* blastoconidia and hyphae. Infection and immunity. 2005; 73:7458–64. https://doi.org/10.1128/IAI.73.11.7458-7464.2005 PMID: 16239547

53. Lowman DW, Greene RR, Bearden DW, Kruppa MD, Pottier M, Monteiro MA, et al. Novel structural features in *Candida albicans* hyphal glucan provide a basis for differential innate immune recognition of hyphae versus yeast. J. Biol. Chem. 2014; 289(6):3432–43. https://doi.org/10.1074/jbc.M113.529131 PMID: 24344127

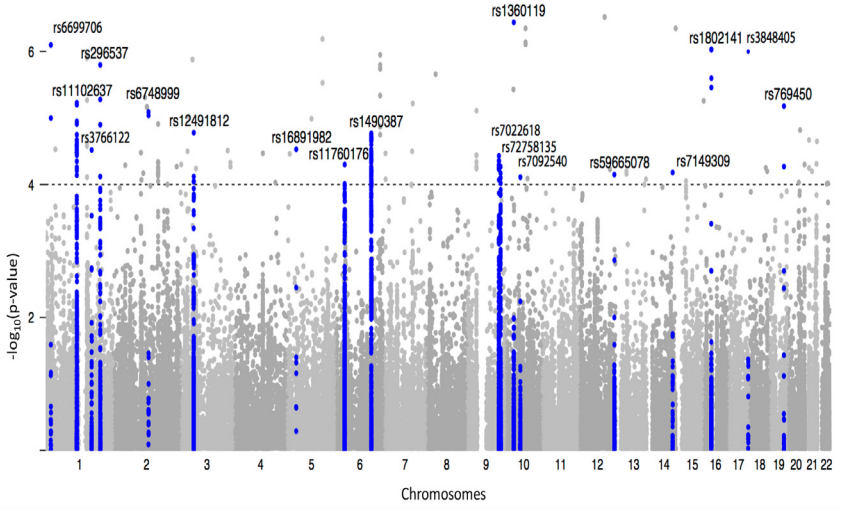
54. Cheng S-C, van de Veerdonk FL, Lenardon M, et al. The dectin-1/inflammasome pathway is responsible for the induction of protective T-helper 17 responses that discriminate between yeasts and hyphae of *Candida albicans*. Journal of Leukocyte Biology. 2011; 90(2):357–366. https://doi.org/10.1189/jlb.1210702 PMID: 21531876

# Supplementary Information

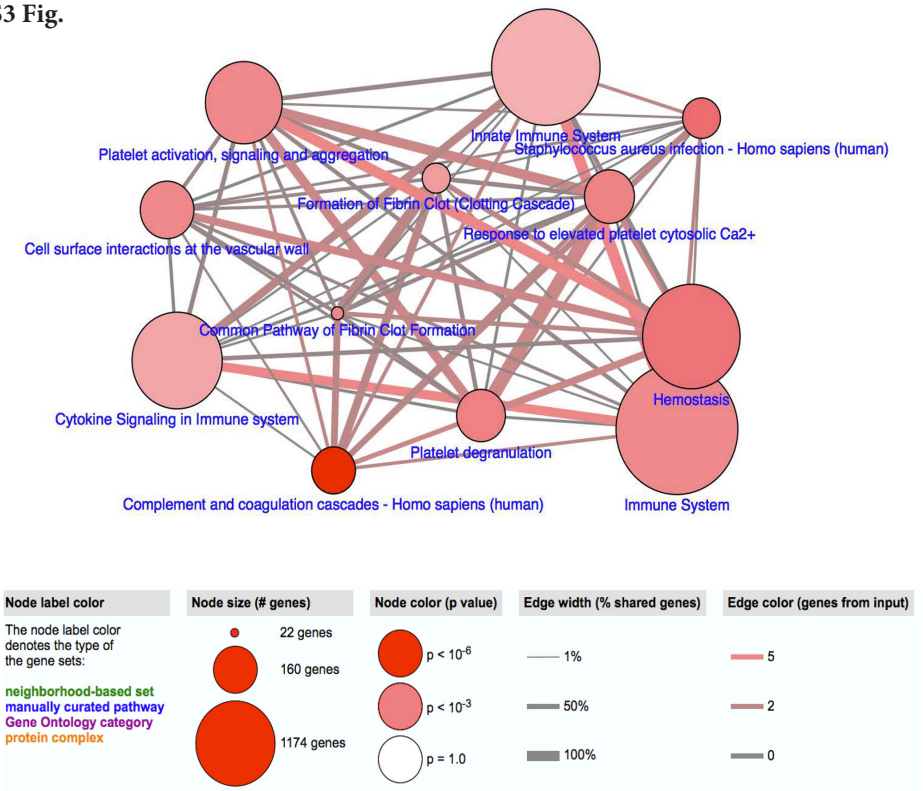
**S1 Fig.**




**S2 Fig.**



**S3 Fig.**


Platelet activation, signaling and aggregation
Innate Immune System
Staphylococcus aureus infection - Homo sapiens (human)
Formation of Fibrin Clot (Clotting Cascade)
Response to elevated platelet cytosolic Ca2+
Cell surface interactions at the vascular wall
Common Pathway of Fibrin Clot Formation
Hemostasis
Cytokine Signaling in Immune system
Platelet degranulation
Complement and coagulation cascades - Homo sapiens (human)
Immune System
Node label color
The node label color denotes the type of the gene sets:
neighborhood-based set
manually curated pathway
Gene Ontology category
protein complex
Node size (# genes)
22 genes
160 genes
1174 genes
Node color (p value)
p < 10-6
p < 10-3
p = 1.0
Edge width (% shared genes)
1%
50%
100%
Edge color (genes from input)
5
2
0

**S4 Fig.**

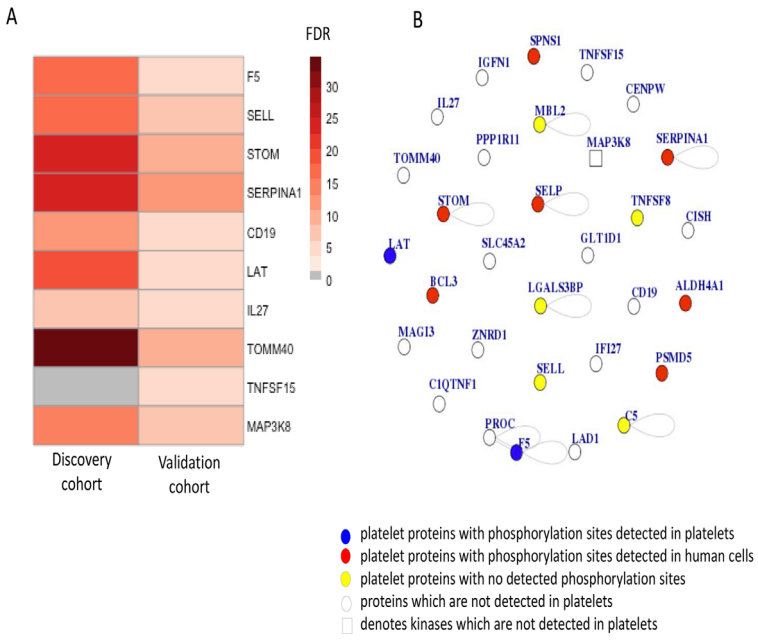

A
FDR
30
25
20
15
10
5
0
F5
SELL
STOM
SERPINA1
CD19
LAT
IL27
TOMM40
TNFSF15
MAP3K8
Discovery cohort
Validation cohort
B
SPNS1
TNFSF15
IGFN1
CENPW
IL27
MBL2
PPP1R11
MAP3K8
SERPINA1
TOMM40
SELP
STOM
TNFSF8
CISH
LAT
SLC45A2
GLT1D1
BCL3
LGALS3BP
CD19
ALDH4A1
MAG13
ZNRD1
IFI27
PSMD5
SELL
C1QTNF1
C5
PROC
F5
LAD1
platelet proteins with phosphorylation sites detected in platelets
platelet proteins with phosphorylation sites detected in human cells
platelet proteins with no detected phosphorylation sites
proteins which are not detected in platelets
denotes kinases which are not detected in platelets

**Table S1.** Differentially-expressed protein-coding genes in response to 4 and 24 hour-*Candida* stimulation that showed >1.5-fold higher expression compared to RPMI medium used as control.

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000114737 | 1.62 | 7.92E-14 | *CISH* | Protein coding | 4 |
| ENSG00000197272 | 2.21 | 1.11E-11 | *IL27* | Protein coding | 4 |
| ENSG00000159166 | 2.32 | 2.02E-11 | *LAD1* | Protein coding | 4 |
| ENSG00000107968 | 1.68 | 1.67E-23 | *MAP3K8* | Protein coding | 4 |
| ENSG00000136244 | 7.22 | 0 | *IL6* | Protein coding | 4 |
| ENSG00000113302 | 6.88 | 4.91E-88 | *IL12B* | Protein coding | 4 |
| ENSG00000115009 | 6.25 | 2.01E-233 | *CCL20* | Protein coding | 4 |
| ENSG00000117525 | 6.19 | 0 | *F3* | Protein coding | 4 |
| ENSG00000006075 | 6.16 | 5.22E-124 | *CCL3* | Protein coding | 4 |
| ENSG00000232810 | 5.95 | 2.66E-110 | *TNF* | Protein coding | 4 |
| ENSG00000115008 | 5.77 | 5.34E-78 | *IL1A* | Protein coding | 4 |
| ENSG00000136688 | 5.21 | 7.96E-51 | *IL36G* | Protein coding | 4 |
| ENSG00000108342 | 5.15 | 2.56E-46 | *CSF3* | Protein coding | 4 |
| ENSG00000129277 | 4.96 | 8.86E-64 | *CCL4* | Protein coding | 4 |
| ENSG00000205021 | 4.77 | 2.68E-39 | *CCL3L1* | Protein coding | 4 |
| ENSG00000125538 | 4.73 | 2.72E-150 | *IL1B* | Protein coding | 4 |
| ENSG00000137869 | 4.73 | 2.22E-43 | *CYP19A1* | Protein coding | 4 |
| ENSG00000205020 | 4.71 | 4.73E-53 | *CCL4L1* | Protein coding | 4 |
| ENSG00000073756 | 4.68 | 9.19E-137 | *PTGS2* | Protein coding | 4 |
| ENSG00000184979 | 4.66 | 1.89E-135 | *USP18* | Protein coding | 4 |
| ENSG00000165474 | 4.56 | 1.47E-40 | *GJB2* | Protein coding | 4 |
| ENSG00000163739 | 4.51 | 1.98E-45 | *CXCL1* | Protein coding | 4 |
| ENSG00000132669 | 4.44 | 7.11E-156 | *RIN2* | Protein coding | 4 |
| ENSG00000050730 | 4.34 | 1.37E-129 | *TNIP3* | Protein coding | 4 |
| ENSG00000197262 | 4.31 | 6.06E-42 | *CCL4L2* | Protein coding | 4 |
| ENSG00000185745 | 4.24 | 1.69E-51 | *IFIT1* | Protein coding | 4 |
| ENSG00000140519 | 4.21 | 1.22E-28 | *RHCG* | Protein coding | 4 |
| ENSG00000134321 | 4.13 | 1.04E-46 | *RSAD2* | Protein coding | 4 |
| ENSG00000123689 | 4.12 | 8.23E-102 | *G0S2* | Protein coding | 4 |
| ENSG00000081041 | 3.90 | 5.76E-37 | *CXCL2* | Protein coding | 4 |
| ENSG00000181634 | 3.89 | 7.72E-72 | *TNFSF15* | Protein coding | 4 |
| ENSG00000177535 | 3.83 | 2.09E-41 | *OR2B11* | Protein coding | 4 |
| ENSG00000108691 | 3.81 | 3.04E-26 | *CCL2* | Protein coding | 4 |
| ENSG00000122641 | 3.81 | 5.40E-29 | *INHBA* | Protein coding | 4 |
| ENSG00000134326 | 3.79 | 3.69E-46 | *CMPK2* | Protein coding | 4 |
| ENSG00000157601 | 3.79 | 3.27E-90 | *MX1* | Protein coding | 4 |
| ENSG00000111537 | 3.68 | 8.05E-21 | *IFNG* | Protein coding | 4 |
| ENSG00000119917 | 3.68 | 6.29E-38 | *IFIT3* | Protein coding | 4 |
| ENSG00000111331 | 3.64 | 6.76E-137 | *OAS3* | Protein coding | 4 |
| ENSG00000137959 | 3.61 | 3.69E-148 | *IFI44L* | Protein coding | 4 |
| ENSG00000154099 | 3.60 | 4.19E-44 | *DNAAF1* | Protein coding | 4 |
| ENSG00000108700 | 3.59 | 4.48E-19 | *CCL8* | Protein coding | 4 |
| ENSG00000169429 | 3.59 | 1.95E-104 | *IL8* | Protein coding | 4 |
| ENSG00000163734 | 3.59 | 3.13E-32 | *CXCL3* | Protein coding | 4 |
| ENSG00000088827 | 3.53 | 6.48E-58 | *SIGLEC1* | Protein coding | 4 |
| ENSG00000078401 | 3.52 | 2.81E-65 | *EDN1* | Protein coding | 4 |
| ENSG00000164400 | 3.48 | 2.09E-24 | *CSF2* | Protein coding | 4 |
| ENSG00000163661 | 3.44 | 4.48E-34 | *PTX3* | Protein coding | 4 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000163735 | 3.42 | 5.16E-36 | *CXCL5* | Protein coding | 4 |
| ENSG00000162711 | 3.41 | 1.02E-293 | *NLRP3* | Protein coding | 4 |
| ENSG00000187608 | 3.40 | 3.22E-30 | *ISG15* | Protein coding | 4 |
| ENSG00000126709 | 3.37 | 6.74E-32 | *IFI6* | Protein coding | 4 |
| ENSG00000123610 | 3.35 | 6.07E-22 | *TNFAIP6* | Protein coding | 4 |
| ENSG00000126262 | 3.34 | 2.85E-25 | *FFAR2* | Protein coding | 4 |
| ENSG00000105855 | 3.32 | 2.42E-55 | *ITGB8* | Protein coding | 4 |
| ENSG00000105707 | 3.29 | 1.01E-23 | *HPN* | Protein coding | 4 |
| ENSG00000167236 | 3.28 | 4.47E-18 | *CCL23* | Protein coding | 4 |
| ENSG00000197632 | 3.26 | 4.45E-90 | *SERPINB2* | Protein coding | 4 |
| ENSG00000157227 | 3.22 | 4.60E-163 | *MMP14* | Protein coding | 4 |
| ENSG00000137965 | 3.19 | 4.36E-66 | *IFI44* | Protein coding | 4 |
| ENSG00000184221 | 3.19 | 1.69E-50 | *OLIG1* | Protein coding | 4 |
| ENSG00000089127 | 3.17 | 2.90E-79 | *OAS1* | Protein coding | 4 |
| ENSG00000123700 | 3.17 | 7.91E-25 | *KCNJ2* | Protein coding | 4 |
| ENSG00000164181 | 3.15 | 3.24E-57 | *ELOVL7* | Protein coding | 4 |
| ENSG00000149289 | 3.08 | 2.73E-65 | *ZC3H12C* | Protein coding | 4 |
| ENSG00000106701 | 3.08 | 6.45E-46 | *FSD1L* | Protein coding | 4 |
| ENSG00000108688 | 3.06 | 3.68E-15 | *CCL7* | Protein coding | 4 |
| ENSG00000078081 | 3.06 | 4.32E-63 | *LAMP3* | Protein coding | 4 |
| ENSG00000125148 | 3.01 | 1.66E-44 | *MT2A* | Protein coding | 4 |
| ENSG00000119922 | 3.00 | 6.85E-22 | *IFIT2* | Protein coding | 4 |
| ENSG00000107104 | 2.99 | 1.01E-96 | *KANK1* | Protein coding | 4 |
| ENSG00000137393 | 2.99 | 5.11E-75 | *RNF144B* | Protein coding | 4 |
| ENSG00000164683 | 2.96 | 2.63E-32 | *HEY1* | Protein coding | 4 |
| ENSG00000136689 | 2.91 | 4.00E-15 | *IL1RN* | Protein coding | 4 |
| ENSG00000138642 | 2.91 | 4.98E-52 | *HERC6* | Protein coding | 4 |
| ENSG00000111335 | 2.91 | 2.31E-66 | *OAS2* | Protein coding | 4 |
| ENSG00000184557 | 2.90 | 2.43E-165 | *SOCS3* | Protein coding | 4 |
| ENSG00000138646 | 2.89 | 2.42E-22 | *HERC5* | Protein coding | 4 |
| ENSG00000102794 | 2.88 | 4.76E-12 | *IRG1* | Protein coding | 4 |
| ENSG00000253831 | 2.81 | 1.08E-13 | *ETV3L* | Protein coding | 4 |
| ENSG00000183486 | 2.79 | 4.07E-55 | *MX2* | Protein coding | 4 |
| ENSG00000121380 | 2.79 | 1.81E-11 | *BCL2L14* | Protein coding | 4 |
| ENSG00000180616 | 2.79 | 6.18E-11 | *SSTR2* | Protein coding | 4 |
| ENSG00000163666 | 2.79 | 1.21E-12 | *HESX1* | Protein coding | 4 |
| ENSG00000162891 | 2.77 | 4.53E-13 | *IL20* | Protein coding | 4 |
| ENSG00000138135 | 2.77 | 2.01E-11 | *CH25H* | Protein coding | 4 |
| ENSG00000136695 | 2.76 | 4.30E-10 | *IL36RN* | Protein coding | 4 |
| ENSG00000105711 | 2.76 | 1.55E-39 | *SCN1B* | Protein coding | 4 |
| ENSG00000114019 | 2.75 | 2.75E-10 | *AMOTL2* | Protein coding | 4 |
| ENSG00000160932 | 2.74 | 9.03E-44 | *LY6E* | Protein coding | 4 |
| ENSG00000132530 | 2.72 | 3.45E-64 | *XAF1* | Protein coding | 4 |
| ENSG00000196141 | 2.69 | 1.97E-35 | *SPATS2L* | Protein coding | 4 |
| ENSG00000144802 | 2.67 | 2.33E-125 | *NFKBIZ* | Protein coding | 4 |
| ENSG00000130589 | 2.66 | 8.17E-68 | *HELZ2* | Protein coding | 4 |
| ENSG00000156127 | 2.64 | 2.69E-44 | *BATF* | Protein coding | 4 |
| ENSG00000104213 | 2.63 | 3.46E-10 | *PDGFRL* | Protein coding | 4 |
| ENSG00000162433 | 2.62 | 7.07E-11 | *AK4* | Protein coding | 4 |
| ENSG00000124875 | 2.62 | 5.85E-14 | *CXCL6* | Protein coding | 4 |
| ENSG00000150782 | 2.60 | 2.80E-42 | *IL18* | Protein coding | 4 |
| ENSG00000173918 | 2.59 | 4.20E-09 | *C1QTNF1* | Protein coding | 4 |
| ENSG00000125462 | 2.59 | 3.66E-30 | *C1orf61* | Protein coding | 4 |
| ENSG00000117009 | 2.57 | 3.41E-36 | *KMO* | Protein coding | 4 |
| ENSG00000107201 | 2.56 | 7.48E-19 | *DDX58* | Protein coding | 4 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000115267 | 2.56 | 8.76E-44 | *IFIH1* | Protein coding | 4 |
| ENSG00000124102 | 2.55 | 3.85E-37 | *PI3* | Protein coding | 4 |
| ENSG00000055332 | 2.55 | 8.28E-56 | *EIF2AK2* | Protein coding | 4 |
| ENSG00000013588 | 2.54 | 5.09E-24 | *GPRC5A* | Protein coding | 4 |
| ENSG00000135373 | 2.54 | 6.60E-17 | *EHF* | Protein coding | 4 |
| ENSG00000171855 | 2.53 | 3.45E-09 | *IFNB1* | Protein coding | 4 |
| ENSG00000177409 | 2.52 | 4.11E-46 | *SAMD9L* | Protein coding | 4 |
| ENSG00000079215 | 2.51 | 3.68E-10 | *SLC1A3* | Protein coding | 4 |
| ENSG00000121900 | 2.51 | 9.54E-11 | *TMEM54* | Protein coding | 4 |
| ENSG00000108753 | 2.50 | 2.70E-09 | *HNF1B* | Protein coding | 4 |
| ENSG00000183657 | 2.50 | 1.10E-09 | *PP13439* | Protein coding | 4 |
| ENSG00000151726 | 2.47 | 1.93E-47 | *ACSL1* | Protein coding | 4 |
| ENSG00000152778 | 2.45 | 9.32E-34 | *IFIT5* | Protein coding | 4 |
| ENSG00000172403 | 2.45 | 2.71E-17 | *SYNPO2* | Protein coding | 4 |
| ENSG00000205927 | 2.41 | 1.36E-14 | *OLIG2* | Protein coding | 4 |
| ENSG00000008118 | 2.39 | 7.85E-11 | *CAMK1G* | Protein coding | 4 |
| ENSG00000163082 | 2.38 | 4.11E-57 | *SGPP2* | Protein coding | 4 |
| ENSG00000128271 | 2.38 | 3.80E-21 | *ADORA2A* | Protein coding | 4 |
| ENSG00000165949 | 2.37 | 1.04E-09 | *IFI27* | Protein coding | 4 |
| ENSG00000140955 | 2.37 | 1.05E-07 | *ADAD2* | Protein coding | 4 |
| ENSG00000187479 | 2.37 | 1.35E-11 | *C11orf96* | Protein coding | 4 |
| ENSG00000172183 | 2.37 | 9.14E-42 | *ISG20* | Protein coding | 4 |
| ENSG00000136048 | 2.36 | 3.60E-33 | *DRAM1* | Protein coding | 4 |
| ENSG00000256515 | 2.36 | 2.43E-07 | *CCL3L3* | Protein coding | 4 |
| ENSG00000133106 | 2.34 | 1.84E-37 | *EPSTI1* | Protein coding | 4 |
| ENSG00000108771 | 2.34 | 2.97E-45 | *DHX58* | Protein coding | 4 |
| ENSG00000068079 | 2.34 | 1.28E-30 | *IFI35* | Protein coding | 4 |
| ENSG00000141837 | 2.34 | 4.49E-19 | *CACNA1A* | Protein coding | 4 |
| ENSG00000134716 | 2.33 | 8.38E-09 | *CYP2J2* | Protein coding | 4 |
| ENSG00000198848 | 2.33 | 3.10E-30 | *CES1* | Protein coding | 4 |
| ENSG00000121743 | 2.31 | 3.10E-16 | *GJA3* | Protein coding | 4 |
| ENSG00000164266 | 2.30 | 2.63E-12 | *SPINK1* | Protein coding | 4 |
| ENSG00000136634 | 2.29 | 2.03E-22 | *IL10* | Protein coding | 4 |
| ENSG00000041982 | 2.28 | 5.04E-08 | *TNC* | Protein coding | 4 |
| ENSG00000136514 | 2.28 | 5.12E-11 | *RTP4* | Protein coding | 4 |
| ENSG00000138821 | 2.27 | 8.87E-15 | *SLC39A8* | Protein coding | 4 |
| ENSG00000112096 | 2.26 | 1.94E-45 | *SOD2* | Protein coding | 4 |
| ENSG00000255398 | 2.26 | 3.27E-24 | *HCAR3* | Protein coding | 4 |
| ENSG00000140379 | 2.26 | 2.16E-32 | *BCL2A1* | Protein coding | 4 |
| ENSG00000196878 | 2.25 | 6.73E-59 | *LAMB3* | Protein coding | 4 |
| ENSG00000196664 | 2.25 | 1.91E-40 | *TLR7* | Protein coding | 4 |
| ENSG00000170312 | 2.24 | 6.28E-19 | *CDK1* | Protein coding | 4 |
| ENSG00000134070 | 2.24 | 3.29E-50 | *IRAK2* | Protein coding | 4 |
| ENSG00000149557 | 2.23 | 8.41E-27 | *FEZ1* | Protein coding | 4 |
| ENSG00000171714 | 2.22 | 1.22E-17 | *ANO5* | Protein coding | 4 |
| ENSG00000105928 | 2.21 | 1.52E-18 | *DFNA5* | Protein coding | 4 |
| ENSG00000139572 | 2.21 | 3.69E-38 | *GPR84* | Protein coding | 4 |
| ENSG00000204960 | 2.21 | 2.45E-07 | *BLACE* | Protein coding | 4 |
| ENSG00000137331 | 2.19 | 4.81E-44 | *IER3* | Protein coding | 4 |
| ENSG00000180875 | 2.18 | 4.98E-12 | *GREM2* | Protein coding | 4 |
| ENSG00000185507 | 2.17 | 7.64E-41 | *IRF7* | Protein coding | 4 |
| ENSG00000102755 | 2.16 | 6.88E-34 | *FLT1* | Protein coding | 4 |
| ENSG00000168389 | 2.16 | 1.15E-41 | *MFSD2A* | Protein coding | 4 |
| ENSG00000122861 | 2.16 | 1.66E-26 | *PLAU* | Protein coding | 4 |
| ENSG00000134470 | 2.15 | 5.61E-25 | *IL15RA* | Protein coding | 4 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000104312 | 2.15 | 2.98E-35 | *RIPK2* | Protein coding | 4 |
| ENSG00000180316 | 2.15 | 1.49E-22 | *PNPLA1* | Protein coding | 4 |
| ENSG00000121858 | 2.15 | 2.37E-11 | *TNFSF10* | Protein coding | 4 |
| ENSG00000157557 | 2.14 | 2.51E-50 | *ETS2* | Protein coding | 4 |
| ENSG00000116690 | 2.13 | 8.94E-21 | *PRG4* | Protein coding | 4 |
| ENSG00000090339 | 2.13 | 1.07E-41 | *ICAM1* | Protein coding | 4 |
| ENSG00000165891 | 2.12 | 1.85E-12 | *E2F7* | Protein coding | 4 |
| ENSG00000138771 | 2.12 | 7.89E-07 | *SHROOM3* | Protein coding | 4 |
| ENSG00000132274 | 2.11 | 7.47E-36 | *TRIM22* | Protein coding | 4 |
| ENSG00000137757 | 2.11 | 6.64E-21 | *CASP5* | Protein coding | 4 |
| ENSG00000137726 | 2.11 | 1.02E-25 | *FXYD6* | Protein coding | 4 |
| ENSG00000155130 | 2.10 | 3.25E-13 | *MARCKS* | Protein coding | 4 |
| ENSG00000143344 | 2.10 | 4.67E-11 | *RGL1* | Protein coding | 4 |
| ENSG00000105825 | 2.09 | 8.31E-15 | *TFPI2* | Protein coding | 4 |
| ENSG00000162614 | 2.09 | 2.37E-14 | *NEXN* | Protein coding | 4 |
| ENSG00000138035 | 2.09 | 7.31E-27 | *PNPT1* | Protein coding | 4 |
| ENSG00000124107 | 2.07 | 2.40E-13 | *SLPI* | Protein coding | 4 |
| ENSG00000010030 | 2.06 | 1.10E-07 | *ETV7* | Protein coding | 4 |
| ENSG00000137628 | 2.05 | 8.04E-28 | *DDX60* | Protein coding | 4 |
| ENSG00000162772 | 2.05 | 3.27E-24 | *ATF3* | Protein coding | 4 |
| ENSG00000182782 | 2.04 | 1.50E-11 | *HCAR2* | Protein coding | 4 |
| ENSG00000110848 | 2.03 | 2.57E-27 | *CD69* | Protein coding | 4 |
| ENSG00000100906 | 2.02 | 6.15E-38 | *NFKBIA* | Protein coding | 4 |
| ENSG00000120217 | 2.01 | 6.95E-10 | *CD274* | Protein coding | 4 |
| ENSG00000114251 | 2.01 | 2.66E-18 | *WNT5A* | Protein coding | 4 |
| ENSG00000178860 | 2.01 | 1.42E-08 | *MSC* | Protein coding | 4 |
| ENSG00000138496 | 2.00 | 5.01E-27 | *PARP9* | Protein coding | 4 |
| ENSG00000115956 | 2.00 | 1.58E-30 | *PLEK* | Protein coding | 4 |
| ENSG00000160326 | 1.99 | 7.69E-33 | *SLC2A6* | Protein coding | 4 |
| ENSG00000185885 | 1.98 | 1.65E-22 | *IFITM1* | Protein coding | 4 |
| ENSG00000131979 | 1.98 | 3.24E-23 | *GCH1* | Protein coding | 4 |
| ENSG00000243649 | 1.98 | 2.32E-07 | *CFB* | Protein coding | 4 |
| ENSG00000142089 | 1.96 | 4.88E-09 | *IFITM3* | Protein coding | 4 |
| ENSG00000121797 | 1.96 | 6.93E-55 | *CCRL2* | Protein coding | 4 |
| ENSG00000166920 | 1.95 | 7.77E-13 | *C15orf48* | Protein coding | 4 |
| ENSG00000026751 | 1.95 | 3.98E-12 | *SLAMF7* | Protein coding | 4 |
| ENSG00000038210 | 1.95 | 1.59E-31 | *PI4K2B* | Protein coding | 4 |
| ENSG00000135114 | 1.93 | 2.13E-23 | *OASL* | Protein coding | 4 |
| ENSG00000196569 | 1.93 | 4.65E-08 | *LAMA2* | Protein coding | 4 |
| ENSG00000166016 | 1.92 | 8.41E-28 | *ABTB2* | Protein coding | 4 |
| ENSG00000104432 | 1.92 | 5.28E-14 | *IL7* | Protein coding | 4 |
| ENSG00000116833 | 1.92 | 8.71E-06 | *NR5A2* | Protein coding | 4 |
| ENSG00000102096 | 1.90 | 3.77E-44 | *PIM2* | Protein coding | 4 |
| ENSG00000205413 | 1.90 | 3.62E-27 | *SAMD9* | Protein coding | 4 |
| ENSG00000075651 | 1.90 | 2.30E-32 | *PLD1* | Protein coding | 4 |
| ENSG00000140464 | 1.90 | 1.47E-34 | *PML* | Protein coding | 4 |
| ENSG00000148926 | 1.89 | 1.28E-39 | *ADM* | Protein coding | 4 |
| ENSG00000185338 | 1.88 | 4.82E-27 | *SOCS1* | Protein coding | 4 |
| ENSG00000163840 | 1.88 | 2.05E-27 | *DTX3L* | Protein coding | 4 |
| ENSG00000116663 | 1.88 | 1.13E-19 | *FBXO6* | Protein coding | 4 |
| ENSG00000179431 | 1.87 | 7.80E-07 | *FJX1* | Protein coding | 4 |
| ENSG00000117594 | 1.87 | 8.77E-05 | *HSD11B1* | Protein coding | 4 |
| ENSG00000135899 | 1.86 | 9.27E-33 | *SP110* | Protein coding | 4 |
| ENSG00000166394 | 1.86 | 1.05E-12 | *CYB5R2* | Protein coding | 4 |
| ENSG00000168961 | 1.86 | 3.85E-37 | *LGALS9* | Protein coding | 4 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000104320 | 1.86 | 2.34E-34 | *NBN* | Protein coding | 4 |
| ENSG00000168062 | 1.85 | 1.12E-05 | *BATF2* | Protein coding | 4 |
| ENSG00000164825 | 1.85 | 2.67E-08 | *DEFB1* | Protein coding | 4 |
| ENSG00000104951 | 1.85 | 2.28E-25 | *IL4I1* | Protein coding | 4 |
| ENSG00000178965 | 1.85 | 6.39E-05 | *C1orf173* | Protein coding | 4 |
| ENSG00000166523 | 1.85 | 4.52E-23 | *CLEC4E* | Protein coding | 4 |
| ENSG00000122643 | 1.84 | 3.72E-20 | *NT5C3* | Protein coding | 4 |
| ENSG00000089692 | 1.83 | 1.09E-22 | *LAG3* | Protein coding | 4 |
| ENSG00000196116 | 1.82 | 2.08E-22 | *TDRD7* | Protein coding | 4 |
| ENSG00000152503 | 1.82 | 1.06E-42 | *TRIM36* | Protein coding | 4 |
| ENSG00000213689 | 1.82 | 2.26E-33 | *TREX1* | Protein coding | 4 |
| ENSG00000059378 | 1.82 | 1.53E-28 | *PARP12* | Protein coding | 4 |
| ENSG00000169194 | 1.80 | 3.90E-06 | *IL13* | Protein coding | 4 |
| ENSG00000142224 | 1.80 | 2.55E-09 | *IL19* | Protein coding | 4 |
| ENSG00000153132 | 1.79 | 1.52E-08 | *CLGN* | Protein coding | 4 |
| ENSG00000197982 | 1.79 | 5.59E-21 | *C1orf122* | Protein coding | 4 |
| ENSG00000164761 | 1.79 | 0.0002 | *TNFRSF11B* | Protein coding | 4 |
| ENSG00000124256 | 1.78 | 1.27E-22 | *ZBP1* | Protein coding | 4 |
| ENSG00000049130 | 1.78 | 3.69E-06 | *KITLG* | Protein coding | 4 |
| ENSG00000140961 | 1.78 | 7.92E-12 | *OSGIN1* | Protein coding | 4 |
| ENSG00000011422 | 1.77 | 1.99E-31 | *PLAUR* | Protein coding | 4 |
| ENSG00000138166 | 1.76 | 3.30E-24 | *DUSP5* | Protein coding | 4 |
| ENSG00000140450 | 1.76 | 2.34E-19 | *ARRDC4* | Protein coding | 4 |
| ENSG00000183473 | 1.76 | 3.58E-16 | *SSTR3* | Protein coding | 4 |
| ENSG00000128383 | 1.75 | 1.13E-07 | *APOBEC3A* | Protein coding | 4 |
| ENSG00000214872 | 1.75 | 3.87E-15 | *SMTNL1* | Protein coding | 4 |
| ENSG00000182168 | 1.75 | 0.00029 | *UNC5C* | Protein coding | 4 |
| ENSG00000169504 | 1.75 | 5.05E-30 | *CLIC4* | Protein coding | 4 |
| ENSG00000188313 | 1.74 | 7.93E-30 | *PLSCR1* | Protein coding | 4 |
| ENSG00000185022 | 1.74 | 4.76E-29 | *MAFF* | Protein coding | 4 |
| ENSG00000166670 | 1.74 | 2.40E-05 | *MMP10* | Protein coding | 4 |
| ENSG00000056558 | 1.73 | 6.74E-58 | *TRAF1* | Protein coding | 4 |
| ENSG00000163565 | 1.72 | 8.34E-24 | *IFI16* | Protein coding | 4 |
| ENSG00000087074 | 1.72 | 7.32E-40 | *PPP1R15A* | Protein coding | 4 |
| ENSG00000183742 | 1.72 | 1.72E-15 | *MACC1* | Protein coding | 4 |
| ENSG00000105246 | 1.71 | 4.87E-17 | *EBI3* | Protein coding | 4 |
| ENSG00000149798 | 1.71 | 1.05E-23 | *CDC42EP2* | Protein coding | 4 |
| ENSG00000136810 | 1.71 | 1.08E-17 | *TXN* | Protein coding | 4 |
| ENSG00000186583 | 1.70 | 2.49E-09 | *SPATC1* | Protein coding | 4 |
| ENSG00000215306 | 1.70 | 0.0002 | *AL135998.1* | Protein coding | 4 |
| ENSG00000221949 | 1.70 | 2.01E-07 | *C12orf61* | Protein coding | 4 |
| ENSG00000125384 | 1.70 | 3.21E-16 | *PTGER2* | Protein coding | 4 |
| ENSG00000130303 | 1.70 | 3.55E-23 | *BST2* | Protein coding | 4 |
| ENSG00000134460 | 1.69 | 3.73E-20 | *IL2RA* | Protein coding | 4 |
| ENSG00000185215 | 1.67 | 6.66E-22 | *TNFAIP2* | Protein coding | 4 |
| ENSG00000075618 | 1.67 | 4.01E-23 | *FSCN1* | Protein coding | 4 |
| ENSG00000123685 | 1.66 | 3.51E-07 | *BATF3* | Protein coding | 4 |
| ENSG00000100024 | 1.65 | 1.15E-11 | *UPB1* | Protein coding | 4 |
| ENSG00000103569 | 1.65 | 2.01E-36 | *AQP9* | Protein coding | 4 |
| ENSG00000118194 | 1.65 | 0.00025 | *TNNT2* | Protein coding | 4 |
| ENSG00000154262 | 1.64 | 1.44E-07 | *ABCA6* | Protein coding | 4 |
| ENSG00000173786 | 1.64 | 1.10E-26 | *CNP* | Protein coding | 4 |
| ENSG00000196684 | 1.64 | 1.10E-20 | *HSH2D* | Protein coding | 4 |
| ENSG00000217825 | 1.64 | 6.04E-05 | *AC099552.4* | Protein coding | 4 |
| ENSG00000143061 | 1.64 | 1.39E-07 | *IGSF3* | Protein coding | 4 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000101438 | 1.64 | 0.0008 | *SLC32A1* | Protein coding | 4 |
| ENSG00000132256 | 1.63 | 4.82E-27 | *TRIM5* | Protein coding | 4 |
| ENSG00000138670 | 1.62 | 5.94E-24 | *RASGEF1B* | Protein coding | 4 |
| ENSG00000029559 | 1.62 | 0.00089 | *IBSP* | Protein coding | 4 |
| ENSG00000181649 | 1.61 | 4.80E-07 | *PHLDA2* | Protein coding | 4 |
| ENSG00000147434 | 1.60 | 1.00E-05 | *CHRNA6* | Protein coding | 4 |
| ENSG00000170525 | 1.60 | 4.70E-35 | *PFKFB3* | Protein coding | 4 |
| ENSG00000112773 | 1.59 | 1.35E-20 | *FAM46A* | Protein coding | 4 |
| ENSG00000137462 | 1.59 | 5.75E-26 | *TLR2* | Protein coding | 4 |
| ENSG00000198673 | 1.59 | 7.92E-14 | *FAM19A2* | Protein coding | 4 |
| ENSG00000156587 | 1.59 | 1.21E-15 | *UBE2L6* | Protein coding | 4 |
| ENSG00000180730 | 1.59 | 4.73E-06 | *SHISA2* | Protein coding | 4 |
| ENSG00000130222 | 1.58 | 2.67E-07 | *GADD45G* | Protein coding | 4 |
| ENSG00000152229 | 1.58 | 1.67E-07 | *PSTPIP2* | Protein coding | 4 |
| ENSG00000171631 | 1.58 | 4.10E-23 | *P2RY6* | Protein coding | 4 |
| ENSG00000196639 | 1.58 | 5.85E-16 | *HRH1* | Protein coding | 4 |
| ENSG00000132109 | 1.58 | 1.75E-15 | *TRIM21* | Protein coding | 4 |
| ENSG00000198121 | 1.56 | 6.69E-08 | *LPAR1* | Protein coding | 4 |
| ENSG00000235531 | 1.56 | 9.37E-08 | *RP11-383H13.1* | Protein coding | 4 |
| ENSG00000169245 | 1.56 | 0.00133 | *CXCL10* | Protein coding | 4 |
| ENSG00000121594 | 1.55 | 3.10E-10 | *CD80* | Protein coding | 4 |
| ENSG00000124882 | 1.55 | 9.77E-24 | *EREG* | Protein coding | 4 |
| ENSG00000148737 | 1.55 | 5.90E-25 | *TCF7L2* | Protein coding | 4 |
| ENSG00000125733 | 1.54 | 7.36E-31 | *TRIP10* | Protein coding | 4 |
| ENSG00000102554 | 1.54 | 4.25E-16 | *KLF5* | Protein coding | 4 |
| ENSG00000125730 | 1.54 | 3.07E-17 | *C3* | Protein coding | 4 |
| ENSG00000135678 | 1.54 | 1.09E-12 | *CPM* | Protein coding | 4 |
| ENSG00000026103 | 1.54 | 1.28E-19 | *FAS* | Protein coding | 4 |
| ENSG00000064787 | 1.54 | 1.03E-06 | *BCAS1* | Protein coding | 4 |
| ENSG00000204099 | 1.53 | 5.39E-16 | *NEU4* | Protein coding | 4 |
| ENSG00000162493 | 1.53 | 1.66E-07 | *PDPN* | Protein coding | 4 |
| ENSG00000130164 | 1.53 | 2.25E-31 | *LDLR* | Protein coding | 4 |
| ENSG00000163644 | 1.53 | 3.36E-23 | *PPM1K* | Protein coding | 4 |
| ENSG00000175471 | 1.53 | 6.86E-18 | *MCTP1* | Protein coding | 4 |
| ENSG00000168334 | 1.53 | 0.00147 | *XIRP1* | Protein coding | 4 |
| ENSG00000167191 | 1.52 | 0.00217 | *GPRC5B* | Protein coding | 4 |
| ENSG00000173801 | 1.52 | 3.42E-19 | *JUP* | Protein coding | 4 |
| ENSG00000182175 | 1.52 | 2.57E-06 | *RGMA* | Protein coding | 4 |
| ENSG00000101331 | 1.52 | 0.00094 | *CCM2L* | Protein coding | 4 |
| ENSG00000188211 | 1.52 | 1.04E-11 | *NCR3LG1* | Protein coding | 4 |
| ENSG00000158470 | 1.52 | 5.45E-30 | *B4GALT5* | Protein coding | 4 |
| ENSG00000074416 | 1.51 | 4.83E-22 | *MGLL* | Protein coding | 4 |
| ENSG00000131669 | 1.51 | 2.90E-30 | *NINJ1* | Protein coding | 4 |
| ENSG00000110047 | 1.50 | 9.54E-49 | *EHD1* | Protein coding | 4 |
| ENSG00000102794 | 8.92 | 0 | *IRG1* | Protein coding | 24 |
| ENSG00000196932 | 7.81 | 1.07E-63 | *TMEM26* | Protein coding | 24 |
| ENSG00000108691 | 7.76 | 3.21E-102 | *CCL2* | Protein coding | 24 |
| ENSG00000163739 | 7.70 | 0 | *CXCL1* | Protein coding | 24 |
| ENSG00000108688 | 7.52 | 4.40E-49 | *CCL7* | Protein coding | 24 |
| ENSG00000163735 | 7.50 | 1.59E-119 | *CXCL5* | Protein coding | 24 |
| ENSG00000165474 | 7.36 | 2.86E-69 | *GJB2* | Protein coding | 24 |
| ENSG00000172379 | 6.86 | 1.38E-57 | *ARNT2* | Protein coding | 24 |
| ENSG00000108700 | 6.86 | 2.04E-30 | *CCL8* | Protein coding | 24 |
| ENSG00000108342 | 6.74 | 1.82E-47 | *CSF3* | Protein coding | 24 |
| ENSG00000172724 | 6.68 | 8.32E-34 | *CCL19* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000125538 | 6.67 | 4.90E-245 | *IL1B* | Protein coding | 24 |
| ENSG00000166396 | 6.61 | 6.77E-38 | *SERPINB7* | Protein coding | 24 |
| ENSG00000124102 | 6.55 | 2.60E-72 | *PI3* | Protein coding | 24 |
| ENSG00000163734 | 6.49 | 2.07E-273 | *CXCL3* | Protein coding | 24 |
| ENSG00000172551 | 6.44 | 2.21E-44 | *MUCL1* | Protein coding | 24 |
| ENSG00000136244 | 6.40 | 2.40E-60 | *IL6* | Protein coding | 24 |
| ENSG00000173918 | 6.37 | 1.07E-35 | *C1QTNF1* | Protein coding | 24 |
| ENSG00000115414 | -6.36 | 1.57E-77 | *FN1* | Protein coding | 24 |
| ENSG00000136688 | 6.31 | 5.92E-47 | *IL36G* | Protein coding | 24 |
| ENSG00000169429 | 6.13 | 0 | *IL8* | Protein coding | 24 |
| ENSG00000006074 | 6.11 | 4.43E-29 | *CCL18* | Protein coding | 24 |
| ENSG00000128917 | 6.02 | 2.23E-32 | *DLL4* | Protein coding | 24 |
| ENSG00000121743 | 6.00 | 3.20E-87 | *GJA3* | Protein coding | 24 |
| ENSG00000197632 | 5.93 | 9.64E-225 | *SERPINB2* | Protein coding | 24 |
| ENSG00000170866 | 5.85 | 2.25E-51 | *LILRA3* | Protein coding | 24 |
| ENSG00000074410 | 5.83 | 3.41E-39 | *CA12* | Protein coding | 24 |
| ENSG00000081041 | 5.82 | 4.56E-234 | *CXCL2* | Protein coding | 24 |
| ENSG00000157227 | 5.81 | 1.91E-243 | *MMP14* | Protein coding | 24 |
| ENSG00000129538 | -5.76 | 1.90E-36 | *RNASE1* | Protein coding | 24 |
| ENSG00000103888 | 5.74 | 1.39E-65 | *KIAA1199* | Protein coding | 24 |
| ENSG00000162493 | 5.73 | 8.09E-29 | *PDPN* | Protein coding | 24 |
| ENSG00000105855 | 5.73 | 1.29E-55 | *ITGB8* | Protein coding | 24 |
| ENSG00000148344 | 5.68 | 6.61E-27 | *PTGES* | Protein coding | 24 |
| ENSG00000099998 | 5.68 | 2.26E-30 | *GGT5* | Protein coding | 24 |
| ENSG00000253831 | 5.61 | 9.55E-34 | *ETV3L* | Protein coding | 24 |
| ENSG00000183762 | 5.59 | 3.27E-55 | *KREMEN1* | Protein coding | 24 |
| ENSG00000006075 | 5.56 | 7.64E-197 | *CCL3* | Protein coding | 24 |
| ENSG00000124875 | 5.50 | 2.60E-46 | *CXCL6* | Protein coding | 24 |
| ENSG00000123610 | 5.38 | 1.58E-74 | *TNFAIP6* | Protein coding | 24 |
| ENSG00000150510 | 5.38 | 1.36E-38 | *FAM124A* | Protein coding | 24 |
| ENSG00000174705 | 5.33 | 3.37E-32 | *SH3PXD2B* | Protein coding | 24 |
| ENSG00000170323 | -5.32 | 4.53E-26 | *FABP4* | Protein coding | 24 |
| ENSG00000105825 | 5.27 | 3.61E-118 | *TFPI2* | Protein coding | 24 |
| ENSG00000123689 | 5.26 | 2.21E-47 | *G0S2* | Protein coding | 24 |
| ENSG00000135047 | 5.21 | 6.92E-81 | *CTSL1* | Protein coding | 24 |
| ENSG00000164181 | 5.21 | 9.38E-80 | *ELOVL7* | Protein coding | 24 |
| ENSG00000101331 | 5.21 | 1.22E-36 | *CCM2L* | Protein coding | 24 |
| ENSG00000154065 | 5.21 | 2.15E-26 | *ANKRD29* | Protein coding | 24 |
| ENSG00000111537 | 5.18 | 6.10E-25 | *IFNG* | Protein coding | 24 |
| ENSG00000163395 | 5.14 | 1.75E-20 | *IGFN1* | Protein coding | 24 |
| ENSG00000110436 | 5.11 | 7.38E-143 | *SLC1A2* | Protein coding | 24 |
| ENSG00000205927 | 5.10 | 9.90E-35 | *OLIG2* | Protein coding | 24 |
| ENSG00000114251 | 5.10 | 3.44E-24 | *WNT5A* | Protein coding | 24 |
| ENSG00000053747 | 5.06 | 2.20E-22 | *LAMA3* | Protein coding | 24 |
| ENSG00000105509 | 5.05 | 2.55E-26 | *HAS1* | Protein coding | 24 |
| ENSG00000073756 | 5.03 | 1.17E-134 | *PTGS2* | Protein coding | 24 |
| ENSG00000105976 | 4.97 | 8.50E-38 | *MET* | Protein coding | 24 |
| ENSG00000050730 | 4.95 | 2.76E-194 | *TNIP3* | Protein coding | 24 |
| ENSG00000164400 | 4.93 | 1.25E-22 | *CSF2* | Protein coding | 24 |
| ENSG00000122641 | 4.93 | 3.41E-41 | *INHBA* | Protein coding | 24 |
| ENSG00000152766 | 4.92 | 1.13E-21 | *ANKRD22* | Protein coding | 24 |
| ENSG00000131203 | 4.89 | 7.16E-64 | *IDO1* | Protein coding | 24 |
| ENSG00000136634 | 4.87 | 2.31E-88 | *IL10* | Protein coding | 24 |
| ENSG00000162433 | 4.87 | 1.09E-63 | *AK4* | Protein coding | 24 |
| ENSG00000019169 | 4.86 | 5.71E-41 | *MARCO* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000156234 | 4.80 | 3.88E-31 | *CXCL13* | Protein coding | 24 |
| ENSG00000196611 | 4.79 | 9.97E-23 | *MMP1* | Protein coding | 24 |
| ENSG00000162407 | 4.72 | 1.77E-104 | *PPAP2B* | Protein coding | 24 |
| ENSG00000115008 | 4.70 | 6.63E-31 | *IL1A* | Protein coding | 24 |
| ENSG00000126262 | 4.70 | 1.01E-30 | *FFAR2* | Protein coding | 24 |
| ENSG00000139572 | 4.70 | 5.68E-118 | *GPR84* | Protein coding | 24 |
| ENSG00000008118 | 4.69 | 1.69E-15 | *CAMK1G* | Protein coding | 24 |
| ENSG00000125730 | 4.69 | 4.24E-47 | *C3* | Protein coding | 24 |
| ENSG00000138685 | 4.62 | 5.42E-19 | *FGF2* | Protein coding | 24 |
| ENSG00000243649 | 4.60 | 9.33E-20 | *CFB* | Protein coding | 24 |
| ENSG00000142224 | 4.58 | 1.12E-25 | *IL19* | Protein coding | 24 |
| ENSG00000121380 | 4.57 | 9.45E-21 | *BCL2L14* | Protein coding | 24 |
| ENSG00000180616 | 4.55 | 2.78E-18 | *SSTR2* | Protein coding | 24 |
| ENSG00000236939 | 4.47 | 1.45E-16 | *C8orf56* | Protein coding | 24 |
| ENSG00000166920 | 4.43 | 1.55E-43 | *C15orf48* | Protein coding | 24 |
| ENSG00000171049 | 4.42 | 1.47E-148 | *FPR2* | Protein coding | 24 |
| ENSG00000164509 | 4.42 | 5.79E-24 | *IL31RA* | Protein coding | 24 |
| ENSG00000213694 | 4.39 | 3.72E-31 | *S1PR3* | Protein coding | 24 |
| ENSG00000137757 | 4.38 | 1.60E-44 | *CASP5* | Protein coding | 24 |
| ENSG00000183019 | 4.33 | 5.15E-75 | *C19orf59* | Protein coding | 24 |
| ENSG00000138135 | 4.33 | 7.79E-22 | *CH25H* | Protein coding | 24 |
| ENSG00000018280 | 4.32 | 2.28E-112 | *SLC11A1* | Protein coding | 24 |
| ENSG00000136695 | 4.31 | 7.88E-15 | *IL36RN* | Protein coding | 24 |
| ENSG00000154188 | -4.25 | 3.30E-27 | *ANGPT1* | Protein coding | 24 |
| ENSG00000124731 | 4.24 | 4.70E-71 | *TREM1* | Protein coding | 24 |
| ENSG00000138316 | 4.24 | 1.92E-54 | *ADAMTS14* | Protein coding | 24 |
| ENSG00000120217 | 4.20 | 4.54E-134 | *CD274* | Protein coding | 24 |
| ENSG00000135111 | 4.18 | 7.62E-21 | *TBX3* | Protein coding | 24 |
| ENSG00000117594 | 4.18 | 1.16E-14 | *HSD11B1* | Protein coding | 24 |
| ENSG00000137673 | 4.18 | 2.91E-36 | *MMP7* | Protein coding | 24 |
| ENSG00000107249 | 4.18 | 2.68E-81 | *GLIS3* | Protein coding | 24 |
| ENSG00000115009 | 4.14 | 8.41E-28 | *CCL20* | Protein coding | 24 |
| ENSG00000105835 | 4.11 | 2.40E-172 | *NAMPT* | Protein coding | 24 |
| ENSG00000184557 | 4.11 | 1.25E-147 | *SOCS3* | Protein coding | 24 |
| ENSG00000229644 | 4.08 | 8.43E-124 | *NAMPTL* | Protein coding | 24 |
| ENSG00000173239 | 4.08 | 7.45E-12 | *LIPM* | Protein coding | 24 |
| ENSG00000136379 | 4.06 | 3.23E-93 | *FAM108C1* | Protein coding | 24 |
| ENSG00000196878 | 4.06 | 3.77E-60 | *LAMB3* | Protein coding | 24 |
| ENSG00000244405 | 4.04 | 3.42E-29 | *ETV5* | Protein coding | 24 |
| ENSG00000198019 | 4.03 | 2.55E-31 | *FCGR1B* | Protein coding | 24 |
| ENSG00000111012 | 4.02 | 6.73E-39 | *CYP27B1* | Protein coding | 24 |
| ENSG00000099250 | 4.01 | 2.31E-115 | *NRP1* | Protein coding | 24 |
| ENSG00000162494 | 4.01 | 1.10E-09 | *LRRC38* | Protein coding | 24 |
| ENSG00000113302 | 4.00 | 1.57E-14 | *IL12B* | Protein coding | 24 |
| ENSG00000187957 | 3.98 | 1.27E-20 | *DNER* | Protein coding | 24 |
| ENSG00000100191 | 3.98 | 2.71E-11 | *SLC5A4* | Protein coding | 24 |
| ENSG00000172594 | 3.97 | 1.53E-152 | *SMPDL3A* | Protein coding | 24 |
| ENSG00000134247 | -3.96 | 4.06E-23 | *PTGFRN* | Protein coding | 24 |
| ENSG00000167236 | 3.95 | 1.20E-25 | *CCL23* | Protein coding | 24 |
| ENSG00000121797 | 3.92 | 1.26E-94 | *CCRL2* | Protein coding | 24 |
| ENSG00000197646 | 3.92 | 2.75E-24 | *PDCD1LG2* | Protein coding | 24 |
| ENSG00000164692 | 3.92 | 1.90E-21 | *COL1A2* | Protein coding | 24 |
| ENSG00000165029 | 3.91 | 9.29E-126 | *ABCA1* | Protein coding | 24 |
| ENSG00000112096 | 3.90 | 1.78E-208 | *SOD2* | Protein coding | 24 |
| ENSG00000133048 | 3.89 | 7.12E-27 | *CHI3L1* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000136048 | 3.89 | 8.49E-98 | *DRAM1* | Protein coding | 24 |
| ENSG00000205021 | 3.88 | 3.46E-24 | *CCL3L1* | Protein coding | 24 |
| ENSG00000196639 | 3.88 | 2.17E-52 | *HRH1* | Protein coding | 24 |
| ENSG00000186818 | 3.87 | 3.82E-20 | *LILRB4* | Protein coding | 24 |
| ENSG00000075651 | 3.84 | 1.10E-115 | *PLD1* | Protein coding | 24 |
| ENSG00000125780 | 3.83 | 5.06E-11 | *TGM3* | Protein coding | 24 |
| ENSG00000138821 | 3.82 | 3.86E-128 | *SLC39A8* | Protein coding | 24 |
| ENSG00000105707 | 3.81 | 1.97E-17 | *HPN* | Protein coding | 24 |
| ENSG00000108702 | 3.81 | 1.44E-08 | *CCL1* | Protein coding | 24 |
| ENSG00000168062 | 3.79 | 9.58E-15 | *BATF2* | Protein coding | 24 |
| ENSG00000145244 | -3.79 | 4.85E-68 | *CORIN* | Protein coding | 24 |
| ENSG00000170525 | 3.79 | 6.75E-145 | *PFKFB3* | Protein coding | 24 |
| ENSG00000157168 | -3.78 | 2.56E-17 | *NRG1* | Protein coding | 24 |
| ENSG00000139567 | 3.78 | 2.50E-17 | *ACVRL1* | Protein coding | 24 |
| ENSG00000148926 | 3.78 | 2.30E-25 | *ADM* | Protein coding | 24 |
| ENSG00000138755 | 3.78 | 2.00E-09 | *CXCL9* | Protein coding | 24 |
| ENSG00000169715 | 3.76 | 2.47E-13 | *MT1E* | Protein coding | 24 |
| ENSG00000100985 | 3.74 | 5.23E-18 | *MMP9* | Protein coding | 24 |
| ENSG00000099985 | 3.74 | 2.55E-63 | *OSM* | Protein coding | 24 |
| ENSG00000103569 | 3.73 | 3.20E-280 | *AQP9* | Protein coding | 24 |
| ENSG00000117228 | 3.72 | 9.18E-30 | *GBP1* | Protein coding | 24 |
| ENSG00000067798 | 3.71 | 2.56E-10 | *NAV3* | Protein coding | 24 |
| ENSG00000169413 | -3.71 | 8.16E-23 | *RNASE6* | Protein coding | 24 |
| ENSG00000164821 | -3.70 | 1.01E-10 | *DEFA4* | Protein coding | 24 |
| ENSG00000253958 | 3.69 | 7.51E-16 | *CLDN23* | Protein coding | 24 |
| ENSG00000150337 | 3.68 | 6.88E-11 | *FCGR1A* | Protein coding | 24 |
| ENSG00000163736 | 3.68 | 4.22E-83 | *PPBP* | Protein coding | 24 |
| ENSG00000169946 | 3.67 | 1.02E-11 | *ZFPM2* | Protein coding | 24 |
| ENSG00000165124 | 3.66 | 5.12E-12 | *SVEP1* | Protein coding | 24 |
| ENSG00000106034 | -3.65 | 3.20E-42 | *CPED1* | Protein coding | 24 |
| ENSG00000133687 | 3.65 | 7.00E-19 | *TMTC1* | Protein coding | 24 |
| ENSG00000128578 | 3.65 | 3.29E-49 | *STRIP2* | Protein coding | 24 |
| ENSG00000010030 | 3.64 | 1.15E-13 | *ETV7* | Protein coding | 24 |
| ENSG00000173369 | 3.63 | 3.79E-14 | *C1QB* | Protein coding | 24 |
| ENSG00000134765 | -3.63 | 3.32E-29 | *DSC1* | Protein coding | 24 |
| ENSG00000105711 | 3.61 | 3.29E-48 | *SCN1B* | Protein coding | 24 |
| ENSG00000122861 | 3.60 | 4.71E-29 | *PLAU* | Protein coding | 24 |
| ENSG00000151726 | 3.60 | 2.57E-219 | *ACSL1* | Protein coding | 24 |
| ENSG00000090659 | 3.60 | 9.77E-29 | *CD209* | Protein coding | 24 |
| ENSG00000197272 | 3.60 | 3.54E-10 | *IL27* | Protein coding | 24 |
| ENSG00000198959 | 3.58 | 4.80E-78 | *TGM2* | Protein coding | 24 |
| ENSG00000140022 | 3.58 | 6.45E-17 | *STON2* | Protein coding | 24 |
| ENSG00000180316 | 3.56 | 5.32E-22 | *PNPLA1* | Protein coding | 24 |
| ENSG00000153208 | 3.56 | 5.94E-20 | *MERTK* | Protein coding | 24 |
| ENSG00000183657 | 3.56 | 2.41E-09 | *PP13439* | Protein coding | 24 |
| ENSG00000134028 | 3.55 | 1.40E-18 | *ADAMDEC1* | Protein coding | 24 |
| ENSG00000198848 | 3.55 | 3.37E-15 | *CES1* | Protein coding | 24 |
| ENSG00000085265 | -3.54 | 1.19E-28 | *FCN1* | Protein coding | 24 |
| ENSG00000164929 | 3.54 | 2.42E-30 | *BAALC* | Protein coding | 24 |
| ENSG00000164266 | 3.54 | 8.39E-17 | *SPINK1* | Protein coding | 24 |
| ENSG00000182580 | -3.54 | 7.82E-10 | *EPHB3* | Protein coding | 24 |
| ENSG00000138944 | 3.53 | 2.80E-10 | *KIAA1644* | Protein coding | 24 |
| ENSG00000149289 | 3.53 | 3.80E-107 | *ZC3H12C* | Protein coding | 24 |
| ENSG00000163518 | 3.53 | 8.66E-21 | *FCRL4* | Protein coding | 24 |
| ENSG00000125144 | 3.52 | 7.82E-08 | *MT1G* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000163661 | 3.52 | 3.13E-46 | *PTX3* | Protein coding | 24 |
| ENSG00000064042 | 3.52 | 1.49E-16 | *LIMCH1* | Protein coding | 24 |
| ENSG00000079215 | 3.51 | 1.05E-13 | *SLC1A3* | Protein coding | 24 |
| ENSG00000189221 | 3.51 | 4.27E-13 | *MAOA* | Protein coding | 24 |
| ENSG00000125148 | 3.50 | 1.93E-21 | *MT2A* | Protein coding | 24 |
| ENSG00000121594 | 3.50 | 2.13E-43 | *CD80* | Protein coding | 24 |
| ENSG00000087076 | 3.49 | 5.59E-20 | *HSD17B14* | Protein coding | 24 |
| ENSG00000164047 | -3.48 | 2.91E-10 | *CAMP* | Protein coding | 24 |
| ENSG00000084734 | 3.48 | 3.23E-08 | *GCKR* | Protein coding | 24 |
| ENSG00000011422 | 3.47 | 3.53E-65 | *PLAUR* | Protein coding | 24 |
| ENSG00000123700 | 3.47 | 2.62E-76 | *KCNJ2* | Protein coding | 24 |
| ENSG00000086300 | 3.46 | 9.51E-138 | *SNX10* | Protein coding | 24 |
| ENSG00000006118 | 3.46 | 2.47E-53 | *TMEM132A* | Protein coding | 24 |
| ENSG00000145685 | 3.45 | 8.19E-69 | *LHFPL2* | Protein coding | 24 |
| ENSG00000105928 | 3.44 | 6.31E-20 | *DFNA5* | Protein coding | 24 |
| ENSG00000134531 | 3.42 | 8.81E-21 | *EMP1* | Protein coding | 24 |
| ENSG00000188820 | 3.40 | 2.88E-12 | *FAM26F* | Protein coding | 24 |
| ENSG00000173372 | 3.36 | 3.63E-14 | *C1QA* | Protein coding | 24 |
| ENSG00000100285 | 3.36 | 1.25E-45 | *NEFH* | Protein coding | 24 |
| ENSG00000112115 | 3.35 | 1.74E-06 | *IL17A* | Protein coding | 24 |
| ENSG00000205856 | 3.35 | 1.68E-06 | *C22orf42* | Protein coding | 24 |
| ENSG00000181634 | 3.34 | 9.00E-34 | *TNFSF15* | Protein coding | 24 |
| ENSG00000161944 | -3.34 | 8.73E-30 | *ASGR2* | Protein coding | 24 |
| ENSG00000174837 | 3.33 | 6.72E-62 | *EMR1* | Protein coding | 24 |
| ENSG00000047936 | 3.33 | 2.09E-06 | *ROS1* | Protein coding | 24 |
| ENSG00000155659 | -3.33 | 6.29E-30 | *VSIG4* | Protein coding | 24 |
| ENSG00000100767 | 3.30 | 1.72E-58 | *PAPLN* | Protein coding | 24 |
| ENSG00000135549 | -3.30 | 4.24E-46 | *PKIB* | Protein coding | 24 |
| ENSG00000100336 | 3.30 | 4.41E-09 | *APOL4* | Protein coding | 24 |
| ENSG00000026751 | 3.30 | 8.28E-62 | *SLAMF7* | Protein coding | 24 |
| ENSG00000130558 | -3.29 | 7.49E-20 | *OLFM1* | Protein coding | 24 |
| ENSG00000149798 | 3.29 | 7.15E-58 | *CDC42EP2* | Protein coding | 24 |
| ENSG00000125810 | 3.29 | 9.66E-59 | *CD93* | Protein coding | 24 |
| ENSG00000172927 | 3.29 | 1.58E-11 | *MYEOV* | Protein coding | 24 |
| ENSG00000172817 | 3.29 | 7.85E-17 | *CYP7B1* | Protein coding | 24 |
| ENSG00000116574 | 3.28 | 2.14E-110 | *RHOU* | Protein coding | 24 |
| ENSG00000127318 | 3.27 | 2.13E-08 | *IL22* | Protein coding | 24 |
| ENSG00000183748 | 3.27 | 2.17E-16 | *MRC1L1* | Protein coding | 24 |
| ENSG00000169908 | 3.26 | 5.95E-09 | *TM4SF1* | Protein coding | 24 |
| ENSG00000178789 | -3.25 | 5.89E-71 | *CD300LB* | Protein coding | 24 |
| ENSG00000181374 | 3.24 | 1.95E-06 | *CCL13* | Protein coding | 24 |
| ENSG00000159450 | 3.24 | 1.24E-09 | *TCHH* | Protein coding | 24 |
| ENSG00000204099 | 3.24 | 3.76E-13 | *NEU4* | Protein coding | 24 |
| ENSG00000132514 | -3.23 | 3.13E-10 | *CLEC10A* | Protein coding | 24 |
| ENSG00000129277 | 3.23 | 3.95E-64 | *CCL4* | Protein coding | 24 |
| ENSG00000128512 | 3.22 | 1.13E-121 | *DOCK4* | Protein coding | 24 |
| ENSG00000172752 | 3.22 | 2.01E-06 | *COL6A5* | Protein coding | 24 |
| ENSG00000124479 | 3.22 | 2.12E-06 | *NDP* | Protein coding | 24 |
| ENSG00000112818 | 3.21 | 5.28E-16 | *MEP1A* | Protein coding | 24 |
| ENSG00000111199 | 3.21 | 2.65E-08 | *TRPV4* | Protein coding | 24 |
| ENSG00000134460 | 3.20 | 3.53E-65 | *IL2RA* | Protein coding | 24 |
| ENSG00000074416 | 3.19 | 7.43E-139 | *MGLL* | Protein coding | 24 |
| ENSG00000101000 | 3.18 | 1.35E-16 | *PROCR* | Protein coding | 24 |
| ENSG00000143344 | 3.18 | 8.58E-15 | *RGL1* | Protein coding | 24 |
| ENSG00000106701 | 3.18 | 9.82E-139 | *FSD1L* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000096006 | -3.18 | 1.31E-11 | *CRISP3* | Protein coding | 24 |
| ENSG00000159189 | 3.17 | 9.07E-12 | *C1QC* | Protein coding | 24 |
| ENSG00000198121 | 3.17 | 3.22E-37 | *LPAR1* | Protein coding | 24 |
| ENSG00000074660 | 3.17 | 2.67E-65 | *SCARF1* | Protein coding | 24 |
| ENSG00000136052 | 3.16 | 1.11E-100 | *SLC41A2* | Protein coding | 24 |
| ENSG00000165238 | 3.16 | 1.74E-14 | *WNK2* | Protein coding | 24 |
| ENSG00000244482 | 3.15 | 2.60E-25 | *LILRA6* | Protein coding | 24 |
| ENSG00000198178 | -3.14 | 2.50E-07 | *CLEC4C* | Protein coding | 24 |
| ENSG00000163823 | 3.14 | 3.93E-20 | *CCR1* | Protein coding | 24 |
| ENSG00000151693 | 3.13 | 3.93E-27 | *ASAP2* | Protein coding | 24 |
| ENSG00000013588 | 3.13 | 1.17E-05 | *GPRC5A* | Protein coding | 24 |
| ENSG00000121900 | 3.13 | 5.88E-09 | *TMEM54* | Protein coding | 24 |
| ENSG00000198829 | 3.12 | 1.22E-10 | *SUCNR1* | Protein coding | 24 |
| ENSG00000180061 | 3.12 | 7.17E-35 | *TMEM150B* | Protein coding | 24 |
| ENSG00000198814 | 3.12 | 4.10E-158 | *GK* | Protein coding | 24 |
| ENSG00000112116 | 3.11 | 1.39E-05 | *IL17F* | Protein coding | 24 |
| ENSG00000060982 | 3.10 | 2.15E-65 | *BCAT1* | Protein coding | 24 |
| ENSG00000175899 | -3.09 | 5.08E-34 | *A2M* | Protein coding | 24 |
| ENSG00000137869 | 3.09 | 2.93E-08 | *CYP19A1* | Protein coding | 24 |
| ENSG00000144681 | 3.08 | 6.11E-10 | *STAC* | Protein coding | 24 |
| ENSG00000104055 | -3.07 | 6.51E-11 | *TGM5* | Protein coding | 24 |
| ENSG00000128342 | 3.07 | 4.20E-22 | *LIF* | Protein coding | 24 |
| ENSG00000154451 | 3.06 | 1.82E-48 | *GBP5* | Protein coding | 24 |
| ENSG00000124490 | -3.06 | 3.60E-06 | *CRISP2* | Protein coding | 24 |
| ENSG00000177575 | 3.06 | 1.03E-17 | *CD163* | Protein coding | 24 |
| ENSG00000052795 | 3.06 | 9.14E-73 | *FNIP2* | Protein coding | 24 |
| ENSG00000151790 | 3.05 | 2.27E-06 | *TDO2* | Protein coding | 24 |
| ENSG00000198535 | 3.05 | 4.84E-08 | *C2CD4A* | Protein coding | 24 |
| ENSG00000165105 | 3.04 | 2.41E-05 | *RASEF* | Protein coding | 24 |
| ENSG00000185338 | 3.04 | 2.83E-42 | *SOCS1* | Protein coding | 24 |
| ENSG00000144837 | 3.03 | 8.83E-09 | *PLA1A* | Protein coding | 24 |
| ENSG00000104951 | 3.03 | 1.39E-10 | *IL4I1* | Protein coding | 24 |
| ENSG00000134256 | -3.03 | 1.33E-46 | *CD101* | Protein coding | 24 |
| ENSG00000134321 | 3.01 | 1.04E-07 | *RSAD2* | Protein coding | 24 |
| ENSG00000205502 | 3.01 | 2.34E-06 | *C2CD4B* | Protein coding | 24 |
| ENSG00000077063 | -3.01 | 9.15E-11 | *CTTNBP2* | Protein coding | 24 |
| ENSG00000180871 | -3.00 | 1.09E-36 | *CXCR2* | Protein coding | 24 |
| ENSG00000178860 | 2.99 | 8.32E-44 | *MSC* | Protein coding | 24 |
| ENSG00000104140 | 2.99 | 2.39E-08 | *RHOV* | Protein coding | 24 |
| ENSG00000147145 | 2.99 | 9.05E-07 | *LPAR4* | Protein coding | 24 |
| ENSG00000111344 | -2.99 | 8.16E-37 | *RASAL1* | Protein coding | 24 |
| ENSG00000198719 | 2.99 | 4.52E-44 | *DLL1* | Protein coding | 24 |
| ENSG00000158428 | 2.98 | 4.96E-10 | *C2orf62* | Protein coding | 24 |
| ENSG00000075420 | 2.98 | 4.75E-128 | *FNDC3B* | Protein coding | 24 |
| ENSG00000121807 | -2.98 | 3.57E-49 | *CCR2* | Protein coding | 24 |
| ENSG00000134532 | 2.97 | 2.91E-29 | *SOX5* | Protein coding | 24 |
| ENSG00000164683 | 2.97 | 3.15E-08 | *HEY1* | Protein coding | 24 |
| ENSG00000259207 | 2.96 | 2.73E-20 | *ITGB3* | Protein coding | 24 |
| ENSG00000143333 | 2.96 | 1.25E-47 | *RGS16* | Protein coding | 24 |
| ENSG00000130052 | 2.95 | 1.95E-48 | *STARD8* | Protein coding | 24 |
| ENSG00000118257 | 2.95 | 8.18E-74 | *NRP2* | Protein coding | 24 |
| ENSG00000134871 | 2.95 | 8.48E-11 | *COL4A2* | Protein coding | 24 |
| ENSG00000104321 | 2.95 | 7.26E-08 | *TRPA1* | Protein coding | 24 |
| ENSG00000109471 | 2.94 | 1.42E-08 | *IL2* | Protein coding | 24 |
| ENSG00000165949 | 2.94 | 1.31E-08 | *IFI27* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000136689 | 2.94 | 7.86E-12 | *IL1RN* | Protein coding | 24 |
| ENSG00000108950 | 2.94 | 4.14E-45 | *FAM20A* | Protein coding | 24 |
| ENSG00000171860 | 2.94 | 7.16E-62 | *C3AR1* | Protein coding | 24 |
| ENSG00000204103 | 2.92 | 3.24E-25 | *MAFB* | Protein coding | 24 |
| ENSG00000243440 | 2.92 | 2.97E-28 | *AF165138.7* | Protein coding | 24 |
| ENSG00000124107 | 2.92 | 5.32E-20 | *SLPI* | Protein coding | 24 |
| ENSG00000095970 | -2.91 | 3.45E-14 | *TREM2* | Protein coding | 24 |
| ENSG00000123342 | 2.90 | 1.30E-18 | *MMP19* | Protein coding | 24 |
| ENSG00000174348 | -2.90 | 1.17E-08 | *PODN* | Protein coding | 24 |
| ENSG00000168952 | -2.88 | 2.69E-07 | *STXBP6* | Protein coding | 24 |
| ENSG00000176170 | 2.88 | 1.67E-41 | *SPHK1* | Protein coding | 24 |
| ENSG00000147650 | 2.88 | 1.80E-51 | *LRP12* | Protein coding | 24 |
| ENSG00000101333 | 2.88 | 4.07E-07 | *PLCB4* | Protein coding | 24 |
| ENSG00000111961 | 2.87 | 5.99E-65 | *SASH1* | Protein coding | 24 |
| ENSG00000133657 | 2.87 | 1.04E-83 | *ATP13A3* | Protein coding | 24 |
| ENSG00000134470 | 2.87 | 1.35E-41 | *IL15RA* | Protein coding | 24 |
| ENSG00000149948 | 2.87 | 4.13E-07 | *HMGA2* | Protein coding | 24 |
| ENSG00000158488 | -2.87 | 1.13E-05 | *CD1E* | Protein coding | 24 |
| ENSG00000139832 | 2.87 | 3.12E-46 | *RAB20* | Protein coding | 24 |
| ENSG00000029153 | 2.87 | 3.77E-60 | *ARNTL2* | Protein coding | 24 |
| ENSG00000174145 | 2.86 | 9.31E-05 | *KIAA1239* | Protein coding | 24 |
| ENSG00000136205 | 2.86 | 3.38E-40 | *TNS3* | Protein coding | 24 |
| ENSG00000148734 | 2.86 | 5.55E-05 | *NPFFR1* | Protein coding | 24 |
| ENSG00000100448 | -2.85 | 0.0001 | *CTSG* | Protein coding | 24 |
| ENSG00000146070 | 2.85 | 2.18E-20 | *PLA2G7* | Protein coding | 24 |
| ENSG00000134326 | 2.85 | 6.78E-08 | *CMPK2* | Protein coding | 24 |
| ENSG00000152503 | 2.84 | 1.84E-26 | *TRIM36* | Protein coding | 24 |
| ENSG00000162512 | 2.84 | 4.60E-14 | *SDC3* | Protein coding | 24 |
| ENSG00000135678 | 2.84 | 4.61E-57 | *CPM* | Protein coding | 24 |
| ENSG00000224821 | 2.84 | 6.70E-06 | *COL4A2-AS2* | Protein coding | 24 |
| ENSG00000115107 | 2.84 | 8.66E-21 | *STEAP3* | Protein coding | 24 |
| ENSG00000166278 | 2.84 | 1.68E-10 | *C2* | Protein coding | 24 |
| ENSG00000175262 | -2.84 | 2.91E-07 | *C1orf127* | Protein coding | 24 |
| ENSG00000137809 | -2.84 | 1.33E-06 | *ITGA11* | Protein coding | 24 |
| ENSG00000105889 | 2.83 | 6.68E-11 | *STEAP1B* | Protein coding | 24 |
| ENSG00000177098 | -2.83 | 2.33E-05 | *SCN4B* | Protein coding | 24 |
| ENSG00000154252 | 2.83 | 9.60E-08 | *GAL3ST2* | Protein coding | 24 |
| ENSG00000129009 | 2.82 | 0.00012 | *ISLR* | Protein coding | 24 |
| ENSG00000186583 | 2.82 | 1.32E-20 | *SPATC1* | Protein coding | 24 |
| ENSG00000188056 | 2.82 | 3.33E-18 | *TREML4* | Protein coding | 24 |
| ENSG00000078098 | 2.81 | 2.96E-07 | *FAP* | Protein coding | 24 |
| ENSG00000101187 | 2.81 | 6.31E-124 | *SLCO4A1* | Protein coding | 24 |
| ENSG00000090339 | 2.80 | 3.16E-45 | *ICAM1* | Protein coding | 24 |
| ENSG00000124491 | -2.80 | 2.00E-23 | *F13A1* | Protein coding | 24 |
| ENSG00000105246 | 2.79 | 1.37E-57 | *EBI3* | Protein coding | 24 |
| ENSG00000176076 | 2.79 | 2.34E-13 | *KCNE1L* | Protein coding | 24 |
| ENSG00000164825 | 2.79 | 8.55E-05 | *DEFB1* | Protein coding | 24 |
| ENSG00000179630 | 2.79 | 1.64E-60 | *LACC1* | Protein coding | 24 |
| ENSG00000184221 | 2.78 | 1.13E-06 | *OLIG1* | Protein coding | 24 |
| ENSG00000151012 | 2.78 | 1.76E-60 | *SLC7A11* | Protein coding | 24 |
| ENSG00000140105 | 2.77 | 1.36E-19 | *WARS* | Protein coding | 24 |
| ENSG00000138061 | 2.77 | 1.30E-63 | *CYP1B1* | Protein coding | 24 |
| ENSG00000197301 | 2.77 | 2.07E-08 | *RP11-366L20.2* | Protein coding | 24 |
| ENSG00000177294 | 2.77 | 3.00E-09 | *FBXO39* | Protein coding | 24 |
| ENSG00000113070 | 2.77 | 2.09E-29 | *HBEGF* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000179331 | 2.76 | 4.55E-28 | *RAB39A* | Protein coding | 24 |
| ENSG00000106341 | 2.76 | 1.79E-18 | *PPP1R17* | Protein coding | 24 |
| ENSG00000169184 | 2.76 | 5.59E-51 | *MN1* | Protein coding | 24 |
| ENSG00000173334 | 2.76 | 8.52E-107 | *TRIB1* | Protein coding | 24 |
| ENSG00000107130 | 2.75 | 1.48E-28 | *NCS1* | Protein coding | 24 |
| ENSG00000116990 | -2.75 | 3.44E-32 | *MYCL1* | Protein coding | 24 |
| ENSG00000119917 | 2.75 | 1.97E-06 | *IFIT3* | Protein coding | 24 |
| ENSG00000140379 | 2.75 | 2.90E-121 | *BCL2A1* | Protein coding | 24 |
| ENSG00000137959 | 2.74 | 2.02E-07 | *IFI44L* | Protein coding | 24 |
| ENSG00000136810 | 2.73 | 2.36E-68 | *TXN* | Protein coding | 24 |
| ENSG00000135424 | 2.73 | 4.37E-14 | *ITGA7* | Protein coding | 24 |
| ENSG00000004846 | 2.73 | 0.00015 | *ABCB5* | Protein coding | 24 |
| ENSG00000158714 | 2.72 | 3.90E-11 | *SLAMF8* | Protein coding | 24 |
| ENSG00000004468 | 2.72 | 3.60E-51 | *CD38* | Protein coding | 24 |
| ENSG00000118322 | 2.72 | 7.21E-14 | *ATP10B* | Protein coding | 24 |
| ENSG00000109099 | 2.72 | 6.38E-14 | *PMP22* | Protein coding | 24 |
| ENSG00000036530 | -2.71 | 9.78E-06 | *CYP46A1* | Protein coding | 24 |
| ENSG00000092200 | -2.71 | 4.18E-13 | *RPGRIP1* | Protein coding | 24 |
| ENSG00000001617 | 2.70 | 1.49E-07 | *SEMA3F* | Protein coding | 24 |
| ENSG00000135821 | 2.70 | 1.03E-40 | *GLUL* | Protein coding | 24 |
| ENSG00000135218 | -2.70 | 6.74E-35 | *CD36* | Protein coding | 24 |
| ENSG00000205364 | 2.69 | 0.00030 | *MT1M* | Protein coding | 24 |
| ENSG00000137462 | 2.69 | 3.92E-51 | *TLR2* | Protein coding | 24 |
| ENSG00000132669 | 2.68 | 4.33E-43 | *RIN2* | Protein coding | 24 |
| ENSG00000073712 | 2.68 | 3.17E-22 | *FERMT2* | Protein coding | 24 |
| ENSG00000108684 | 2.68 | 0.00033 | *ASIC2* | Protein coding | 24 |
| ENSG00000155254 | -2.68 | 1.88E-31 | *MARVELD1* | Protein coding | 24 |
| ENSG00000163814 | 2.67 | 4.69E-42 | *CDCP1* | Protein coding | 24 |
| ENSG00000135636 | 2.67 | 1.96E-11 | *DYSF* | Protein coding | 24 |
| ENSG00000114248 | 2.67 | 2.30E-06 | *LRRC31* | Protein coding | 24 |
| ENSG00000089041 | 2.67 | 1.08E-15 | *P2RX7* | Protein coding | 24 |
| ENSG00000138356 | 2.67 | 0.00022 | *AOX1* | Protein coding | 24 |
| ENSG00000185052 | 2.67 | 1.72E-18 | *SLC24A3* | Protein coding | 24 |
| ENSG00000169245 | 2.67 | 3.21E-05 | *CXCL10* | Protein coding | 24 |
| ENSG00000146374 | 2.67 | 0.00036 | *RSPO3* | Protein coding | 24 |
| ENSG00000155629 | 2.66 | 4.26E-88 | *PIK3AP1* | Protein coding | 24 |
| ENSG00000157827 | 2.65 | 1.47E-69 | *FMNL2* | Protein coding | 24 |
| ENSG00000064787 | 2.65 | 7.68E-23 | *BCAS1* | Protein coding | 24 |
| ENSG00000131979 | 2.65 | 5.59E-91 | *GCH1* | Protein coding | 24 |
| ENSG00000165935 | 2.64 | 3.41E-05 | *C12orf70* | Protein coding | 24 |
| ENSG00000104918 | 2.64 | 3.55E-16 | *RETN* | Protein coding | 24 |
| ENSG00000158473 | -2.64 | 6.85E-24 | *CD1D* | Protein coding | 24 |
| ENSG00000117013 | 2.64 | 5.36E-15 | *KCNQ4* | Protein coding | 24 |
| ENSG00000180113 | 2.64 | 1.28E-57 | *TDRD6* | Protein coding | 24 |
| ENSG00000059728 | 2.64 | 4.41E-68 | *MXD1* | Protein coding | 24 |
| ENSG00000134070 | 2.63 | 7.47E-65 | *IRAK2* | Protein coding | 24 |
| ENSG00000101188 | 2.63 | 4.02E-11 | *NTSR1* | Protein coding | 24 |
| ENSG00000158869 | 2.63 | 2.68E-53 | *FCER1G* | Protein coding | 24 |
| ENSG00000140859 | 2.61 | 1.04E-16 | *KIFC3* | Protein coding | 24 |
| ENSG00000075618 | 2.60 | 6.27E-83 | *FSCN1* | Protein coding | 24 |
| ENSG00000101336 | 2.60 | 2.04E-51 | *HCK* | Protein coding | 24 |
| ENSG00000116663 | 2.60 | 7.68E-12 | *FBXO6* | Protein coding | 24 |
| ENSG00000165621 | -2.60 | 0.00055 | *OXGR1* | Protein coding | 24 |
| ENSG00000171659 | -2.60 | 1.74E-15 | *GPR34* | Protein coding | 24 |
| ENSG00000215306 | 2.59 | 8.20E-07 | *AL135998.1* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000188290 | 2.59 | 0.00022 | *HES4* | Protein coding | 24 |
| ENSG00000197249 | 2.59 | 4.87E-18 | *SERPINA1* | Protein coding | 24 |
| ENSG00000182782 | 2.59 | 1.90E-25 | *HCAR2* | Protein coding | 24 |
| ENSG00000106853 | 2.59 | 1.20E-19 | *PTGR1* | Protein coding | 24 |
| ENSG00000254087 | 2.58 | 2.36E-81 | *LYN* | Protein coding | 24 |
| ENSG00000109265 | 2.58 | 1.62E-12 | *KIAA1211* | Protein coding | 24 |
| ENSG00000075223 | 2.58 | 6.62E-56 | *SEMA3C* | Protein coding | 24 |
| ENSG00000025434 | 2.57 | 7.77E-14 | *NR1H3* | Protein coding | 24 |
| ENSG00000174502 | 2.57 | 6.76E-05 | *SLC26A9* | Protein coding | 24 |
| ENSG00000185897 | 2.57 | 0.00067 | *FFAR3* | Protein coding | 24 |
| ENSG00000197046 | 2.57 | 1.77E-15 | *SIGLEC15* | Protein coding | 24 |
| ENSG00000104972 | 2.56 | 2.07E-37 | *LILRB1* | Protein coding | 24 |
| ENSG00000140279 | 2.56 | 5.68E-20 | *DUOX2* | Protein coding | 24 |
| ENSG00000188313 | 2.56 | 9.63E-19 | *PLSCR1* | Protein coding | 24 |
| ENSG00000100024 | 2.56 | 2.49E-28 | *UPB1* | Protein coding | 24 |
| ENSG00000175352 | 2.56 | 3.08E-25 | *NRIP3* | Protein coding | 24 |
| ENSG00000205020 | 2.55 | 3.61E-07 | *CCL4L1* | Protein coding | 24 |
| ENSG00000148677 | 2.55 | 0.00011 | *ANKRD1* | Protein coding | 24 |
| ENSG00000141161 | -2.55 | 1.22E-05 | *UNC45B* | Protein coding | 24 |
| ENSG00000179431 | 2.55 | 4.78E-06 | *FJX1* | Protein coding | 24 |
| ENSG00000007350 | -2.54 | 2.73E-14 | *TKTL1* | Protein coding | 24 |
| ENSG00000124780 | -2.54 | 0.00077 | *KCNK17* | Protein coding | 24 |
| ENSG00000161381 | -2.54 | 5.76E-22 | *PLXDC1* | Protein coding | 24 |
| ENSG00000123685 | 2.53 | 1.52E-23 | *BATF3* | Protein coding | 24 |
| ENSG00000002587 | 2.53 | 8.85E-07 | *HS3ST1* | Protein coding | 24 |
| ENSG00000268865 | 2.52 | 1.69E-05 | *AC026310.1* | Protein coding | 24 |
| ENSG00000255398 | 2.52 | 1.36E-15 | *HCAR3* | Protein coding | 24 |
| ENSG00000166165 | 2.52 | 4.66E-20 | *CKB* | Protein coding | 24 |
| ENSG00000182566 | 2.52 | 2.73E-06 | *CLEC4G* | Protein coding | 24 |
| ENSG00000147454 | 2.52 | 1.69E-38 | *SLC25A37* | Protein coding | 24 |
| ENSG00000176049 | 2.51 | 4.40E-38 | *JAKMIP2* | Protein coding | 24 |
| ENSG00000163694 | 2.51 | 7.20E-45 | *RBM47* | Protein coding | 24 |
| ENSG00000159399 | 2.51 | 8.62E-41 | *HK2* | Protein coding | 24 |
| ENSG00000235531 | 2.51 | 2.64E-20 | *RP11-383H13.1* | Protein coding | 24 |
| ENSG00000155465 | 2.51 | 2.21E-31 | *SLC7A7* | Protein coding | 24 |
| ENSG00000184979 | 2.51 | 2.30E-06 | *USP18* | Protein coding | 24 |
| ENSG00000130202 | 2.51 | 3.06E-17 | *PVRL2* | Protein coding | 24 |
| ENSG00000152580 | -2.50 | 9.73E-07 | *IGSF10* | Protein coding | 24 |
| ENSG00000077585 | 2.50 | 6.41E-94 | *GPR137B* | Protein coding | 24 |
| ENSG00000112394 | 2.49 | 2.48E-08 | *SLC16A10* | Protein coding | 24 |
| ENSG00000137393 | 2.49 | 7.52E-111 | *RNF144B* | Protein coding | 24 |
| ENSG00000205436 | 2.48 | 0.00025 | *EXOC3L4* | Protein coding | 24 |
| ENSG00000137331 | 2.48 | 7.53E-47 | *IER3* | Protein coding | 24 |
| ENSG00000187116 | 2.48 | 9.38E-11 | *LILRA5* | Protein coding | 24 |
| ENSG00000135094 | 2.48 | 9.63E-07 | *SDS* | Protein coding | 24 |
| ENSG00000002549 | 2.48 | 7.34E-11 | *LAP3* | Protein coding | 24 |
| ENSG00000104093 | 2.47 | 1.76E-61 | *DMXL2* | Protein coding | 24 |
| ENSG00000188676 | 2.47 | 4.30E-24 | *IDO2* | Protein coding | 24 |
| ENSG00000131042 | 2.47 | 1.41E-30 | *LILRB2* | Protein coding | 24 |
| ENSG00000251139 | 2.47 | 0.00118 | *RP11-701P16.2* | Protein coding | 24 |
| ENSG00000228835 | -2.46 | 8.72E-07 | *AC012123.1* | Protein coding | 24 |
| ENSG00000120586 | 2.46 | 2.03E-06 | *MRC1* | Protein coding | 24 |
| ENSG00000163293 | -2.46 | 2.68E-12 | *NIPAL1* | Protein coding | 24 |
| ENSG00000180509 | 2.46 | 2.55E-24 | *KCNE1* | Protein coding | 24 |
| ENSG00000149573 | -2.46 | 3.02E-19 | *MPZL2* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000105419 | 2.45 | 2.88E-05 | *MEIS3* | Protein coding | 24 |
| ENSG00000110079 | 2.45 | 1.15E-09 | *MS4A4A* | Protein coding | 24 |
| ENSG00000131669 | 2.45 | 1.97E-43 | *NINJ1* | Protein coding | 24 |
| ENSG00000178445 | 2.45 | 2.91E-09 | *GLDC* | Protein coding | 24 |
| ENSG00000213512 | 2.45 | 1.07E-17 | *GBP7* | Protein coding | 24 |
| ENSG00000133116 | 2.45 | 3.87E-16 | *KL* | Protein coding | 24 |
| ENSG00000114737 | 2.45 | 2.51E-26 | *CISH* | Protein coding | 24 |
| ENSG00000133106 | 2.44 | 3.12E-09 | *EPSTI1* | Protein coding | 24 |
| ENSG00000122043 | -2.44 | 1.08E-06 | *LINC00544* | Protein coding | 24 |
| ENSG00000120129 | 2.44 | 4.94E-56 | *DUSP1* | Protein coding | 24 |
| ENSG00000157064 | 2.44 | 0.00036 | *NMNAT2* | Protein coding | 24 |
| ENSG00000188786 | 2.44 | 5.91E-71 | *MTF1* | Protein coding | 24 |
| ENSG00000175183 | 2.44 | 0.00064 | *CSRP2* | Protein coding | 24 |
| ENSG00000155130 | 2.43 | 3.27E-64 | *MARCKS* | Protein coding | 24 |
| ENSG00000178726 | 2.43 | 1.10E-21 | *THBD* | Protein coding | 24 |
| ENSG00000115415 | 2.43 | 3.56E-10 | *STAT1* | Protein coding | 24 |
| ENSG00000197122 | 2.43 | 5.18E-33 | *SRC* | Protein coding | 24 |
| ENSG00000104432 | 2.43 | 6.73E-42 | *IL7* | Protein coding | 24 |
| ENSG00000172216 | 2.43 | 3.39E-35 | *CEBPB* | Protein coding | 24 |
| ENSG00000103196 | 2.42 | 3.38E-38 | *CRISPLD2* | Protein coding | 24 |
| ENSG00000123095 | 2.42 | 4.13E-35 | *BHLHE41* | Protein coding | 24 |
| ENSG00000180251 | 2.42 | 0.00013 | *SLC9A4* | Protein coding | 24 |
| ENSG00000162897 | 2.42 | 0.00039 | *FCAMR* | Protein coding | 24 |
| ENSG00000110077 | -2.40 | 8.53E-24 | *MS4A6A* | Protein coding | 24 |
| ENSG00000058085 | 2.40 | 9.64E-06 | *LAMC2* | Protein coding | 24 |
| ENSG00000160593 | -2.40 | 3.55E-58 | *AMICA1* | Protein coding | 24 |
| ENSG00000170448 | -2.39 | 1.05E-111 | *NFXL1* | Protein coding | 24 |
| ENSG00000103313 | 2.39 | 1.09E-28 | *MEFV* | Protein coding | 24 |
| ENSG00000185215 | 2.38 | 9.50E-39 | *TNFAIP2* | Protein coding | 24 |
| ENSG00000100906 | 2.37 | 8.18E-74 | *NFKBIA* | Protein coding | 24 |
| ENSG00000169403 | 2.36 | 1.94E-29 | *PTAFR* | Protein coding | 24 |
| ENSG00000179813 | 2.36 | 0.00125 | *FAM216B* | Protein coding | 24 |
| ENSG00000130066 | 2.36 | 2.54E-59 | *SAT1* | Protein coding | 24 |
| ENSG00000077327 | 2.36 | 0.00219 | *SPAG6* | Protein coding | 24 |
| ENSG00000101460 | 2.35 | 3.60E-17 | *MAP1LC3A* | Protein coding | 24 |
| ENSG00000078114 | -2.35 | 7.03E-05 | *NEBL* | Protein coding | 24 |
| ENSG00000184500 | 2.34 | 1.29E-05 | *PROS1* | Protein coding | 24 |
| ENSG00000113916 | 2.34 | 3.08E-131 | *BCL6* | Protein coding | 24 |
| ENSG00000185499 | 2.34 | 3.10E-15 | *MUC1* | Protein coding | 24 |
| ENSG00000162654 | 2.34 | 2.82E-37 | *GBP4* | Protein coding | 24 |
| ENSG00000088826 | 2.33 | 2.64E-23 | *SMOX* | Protein coding | 24 |
| ENSG00000139318 | 2.33 | 5.15E-33 | *DUSP6* | Protein coding | 24 |
| ENSG00000170961 | 2.33 | 0.00255 | *HAS2* | Protein coding | 24 |
| ENSG00000187796 | -2.33 | 3.38E-39 | *CARD9* | Protein coding | 24 |
| ENSG00000158470 | 2.33 | 2.83E-52 | *B4GALT5* | Protein coding | 24 |
| ENSG00000153246 | 2.32 | 0.00025 | *PLA2R1* | Protein coding | 24 |
| ENSG00000176788 | 2.32 | 3.11E-65 | *BASP1* | Protein coding | 24 |
| ENSG00000163220 | 2.32 | 6.47E-10 | *S100A9* | Protein coding | 24 |
| ENSG00000186907 | 2.31 | 1.10E-06 | *RTN4RL2* | Protein coding | 24 |
| ENSG00000095203 | -2.31 | 0.00015 | *EPB41L4B* | Protein coding | 24 |
| ENSG00000171596 | -2.31 | 2.16E-17 | *NMUR1* | Protein coding | 24 |
| ENSG00000173805 | -2.31 | 6.24E-05 | *HAP1* | Protein coding | 24 |
| ENSG00000143067 | 2.31 | 4.66E-46 | *ZNF697* | Protein coding | 24 |
| ENSG00000254521 | 2.30 | 2.67E-07 | *SIGLEC12* | Protein coding | 24 |
| ENSG00000159713 | -2.30 | 0.00131 | *TPPP3* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000078401 | 2.30 | 2.73E-24 | *EDN1* | Protein coding | 24 |
| ENSG00000164733 | 2.30 | 1.76E-29 | *CTSB* | Protein coding | 24 |
| ENSG00000186431 | 2.30 | 1.52E-25 | *FCAR* | Protein coding | 24 |
| ENSG00000107562 | 2.29 | 0.0013 | *CXCL12* | Protein coding | 24 |
| ENSG00000135373 | 2.29 | 1.15E-16 | *EHF* | Protein coding | 24 |
| ENSG00000168615 | 2.29 | 8.33E-67 | *ADAM9* | Protein coding | 24 |
| ENSG00000134802 | 2.29 | 1.64E-44 | *SLC43A3* | Protein coding | 24 |
| ENSG00000197208 | 2.29 | 2.89E-22 | *SLC22A4* | Protein coding | 24 |
| ENSG00000042493 | 2.29 | 3.98E-16 | *CAPG* | Protein coding | 24 |
| ENSG00000196839 | 2.29 | 2.01E-58 | *ADA* | Protein coding | 24 |
| ENSG00000120833 | 2.28 | 1.21E-17 | *SOCS2* | Protein coding | 24 |
| ENSG00000177409 | 2.28 | 1.63E-12 | *SAMD9L* | Protein coding | 24 |
| ENSG00000196935 | 2.28 | 4.37E-18 | *SRGAP1* | Protein coding | 24 |
| ENSG00000181577 | 2.28 | 1.28E-12 | *C6orf223* | Protein coding | 24 |
| ENSG00000198417 | 2.28 | 8.81E-33 | *MT1F* | Protein coding | 24 |
| ENSG00000134853 | 2.27 | 0.00040 | *PDGFRA* | Protein coding | 24 |
| ENSG00000187678 | 2.27 | 0.00113 | *SPRY4* | Protein coding | 24 |
| ENSG00000008394 | 2.26 | 2.25E-24 | *MGST1* | Protein coding | 24 |
| ENSG00000138496 | 2.26 | 2.32E-11 | *PARP9* | Protein coding | 24 |
| ENSG00000024422 | 2.26 | 9.46E-05 | *EHD2* | Protein coding | 24 |
| ENSG00000114315 | 2.25 | 1.25E-07 | *HES1* | Protein coding | 24 |
| ENSG00000171989 | 2.25 | 1.16E-05 | *LDHAL6B* | Protein coding | 24 |
| ENSG00000154856 | 2.25 | 2.79E-06 | *APCDD1* | Protein coding | 24 |
| ENSG00000137965 | 2.25 | 9.13E-07 | *IFI44* | Protein coding | 24 |
| ENSG00000138080 | 2.25 | 2.57E-16 | *EMILIN1* | Protein coding | 24 |
| ENSG00000183347 | 2.25 | 4.52E-11 | *GBP6* | Protein coding | 24 |
| ENSG00000168334 | 2.24 | 7.97E-06 | *XIRP1* | Protein coding | 24 |
| ENSG00000170681 | 2.24 | 4.87E-08 | *MURC* | Protein coding | 24 |
| ENSG00000128383 | 2.24 | 2.72E-05 | *APOBEC3A* | Protein coding | 24 |
| ENSG00000171657 | -2.24 | 1.19E-08 | *GPR82* | Protein coding | 24 |
| ENSG00000153132 | 2.24 | 1.88E-05 | *CLGN* | Protein coding | 24 |
| ENSG00000185686 | 2.24 | 0.00153 | *PRAME* | Protein coding | 24 |
| ENSG00000108244 | 2.23 | 4.32E-06 | *KRT23* | Protein coding | 24 |
| ENSG00000128641 | 2.23 | 5.43E-18 | *MYO1B* | Protein coding | 24 |
| ENSG00000078081 | 2.23 | 6.88E-48 | *LAMP3* | Protein coding | 24 |
| ENSG00000136002 | -2.23 | 2.91E-07 | *ARHGEF4* | Protein coding | 24 |
| ENSG00000115919 | 2.23 | 2.65E-76 | *KYNU* | Protein coding | 24 |
| ENSG00000152229 | 2.23 | 2.00E-37 | *PSTPIP2* | Protein coding | 24 |
| ENSG00000186354 | 2.23 | 0.00178 | *C9orf47* | Protein coding | 24 |
| ENSG00000111331 | 2.22 | 1.17E-05 | *OAS3* | Protein coding | 24 |
| ENSG00000106066 | -2.22 | 3.54E-21 | *CPVL* | Protein coding | 24 |
| ENSG00000113763 | 2.22 | 0.00089 | *UNC5A* | Protein coding | 24 |
| ENSG00000124145 | 2.22 | 9.04E-37 | *SDC4* | Protein coding | 24 |
| ENSG00000157766 | 2.22 | 0.00040 | *ACAN* | Protein coding | 24 |
| ENSG00000169862 | 2.22 | 0.00105 | *CTNND2* | Protein coding | 24 |
| ENSG00000136026 | 2.22 | 2.98E-37 | *CKAP4* | Protein coding | 24 |
| ENSG00000189325 | 2.21 | 0.00074 | *C6orf222* | Protein coding | 24 |
| ENSG00000265096 | 2.21 | 0.00448 | *AC073624.1* | Protein coding | 24 |
| ENSG00000138119 | 2.21 | 1.26E-08 | *MYOF* | Protein coding | 24 |
| ENSG00000169248 | 2.21 | 0.00016 | *CXCL11* | Protein coding | 24 |
| ENSG00000254827 | 2.21 | 1.47E-11 | *SLC22A18AS* | Protein coding | 24 |
| ENSG00000254415 | 2.20 | 7.44E-27 | *SIGLEC14* | Protein coding | 24 |
| ENSG00000156219 | 2.20 | 0.00403 | *ART3* | Protein coding | 24 |
| ENSG00000198682 | 2.20 | 1.75E-33 | *PAPSS2* | Protein coding | 24 |
| ENSG00000221887 | 2.20 | 5.19E-05 | *HMSD* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000175857 | -2.19 | 1.44E-29 | *GAPT* | Protein coding | 24 |
| ENSG00000165914 | 2.19 | 2.37E-17 | *TTC7B* | Protein coding | 24 |
| ENSG00000124762 | 2.19 | 2.97E-30 | *CDKN1A* | Protein coding | 24 |
| ENSG00000176046 | 2.19 | 0.00305 | *NUPR1* | Protein coding | 24 |
| ENSG00000104368 | 2.19 | 4.30E-19 | *PLAT* | Protein coding | 24 |
| ENSG00000136514 | 2.18 | 7.09E-17 | *RTP4* | Protein coding | 24 |
| ENSG00000170486 | -2.18 | 3.53E-45 | *KRT72* | Protein coding | 24 |
| ENSG00000214274 | -2.18 | 1.53E-08 | *ANG* | Protein coding | 24 |
| ENSG00000104312 | 2.18 | 7.30E-45 | *RIPK2* | Protein coding | 24 |
| ENSG00000213886 | 2.18 | 0.00531 | *UBD* | Protein coding | 24 |
| ENSG00000162745 | 2.18 | 1.47E-06 | *OLFML2B* | Protein coding | 24 |
| ENSG00000221963 | 2.17 | 8.10E-41 | *APOL6* | Protein coding | 24 |
| ENSG00000149635 | 2.17 | 2.33E-08 | *OCSTAMP* | Protein coding | 24 |
| ENSG00000095383 | 2.17 | 8.46E-22 | *TBC1D2* | Protein coding | 24 |
| ENSG00000130368 | 2.17 | 0.00551 | *MAS1* | Protein coding | 24 |
| ENSG00000116514 | 2.17 | 1.64E-43 | *RNF19B* | Protein coding | 24 |
| ENSG00000149927 | -2.17 | 0.00075 | *DOC2A* | Protein coding | 24 |
| ENSG00000225485 | 2.16 | 4.92E-09 | *ARHGAP23* | Protein coding | 24 |
| ENSG00000132530 | 2.16 | 9.92E-08 | *XAF1* | Protein coding | 24 |
| ENSG00000143545 | 2.16 | 1.18E-26 | *RAB13* | Protein coding | 24 |
| ENSG00000117009 | 2.16 | 2.93E-39 | *KMO* | Protein coding | 24 |
| ENSG00000042062 | 2.16 | 1.23E-20 | *FAM65C* | Protein coding | 24 |
| ENSG00000136960 | 2.16 | 2.44E-22 | *ENPP2* | Protein coding | 24 |
| ENSG00000167996 | 2.16 | 1.32E-28 | *FTH1* | Protein coding | 24 |
| ENSG00000166068 | 2.15 | 4.71E-23 | *SPRED1* | Protein coding | 24 |
| ENSG00000154027 | -2.15 | 2.57E-14 | *AK5* | Protein coding | 24 |
| ENSG00000149557 | 2.15 | 1.82E-23 | *FEZ1* | Protein coding | 24 |
| ENSG00000198369 | 2.15 | 1.05E-25 | *SPRED2* | Protein coding | 24 |
| ENSG00000111452 | -2.15 | 1.37E-14 | *GPR133* | Protein coding | 24 |
| ENSG00000120306 | 2.14 | 1.51E-26 | *CYSTM1* | Protein coding | 24 |
| ENSG00000149131 | 2.14 | 0.00145 | *SERPING1* | Protein coding | 24 |
| ENSG00000142185 | 2.13 | 7.82E-12 | *TRPM2* | Protein coding | 24 |
| ENSG00000163464 | -2.13 | 2.58E-12 | *CXCR1* | Protein coding | 24 |
| ENSG00000232810 | 2.13 | 6.46E-43 | *TNF* | Protein coding | 24 |
| ENSG00000138166 | 2.13 | 2.87E-59 | *DUSP5* | Protein coding | 24 |
| ENSG00000185022 | 2.13 | 3.07E-34 | *MAFF* | Protein coding | 24 |
| ENSG00000143546 | 2.13 | 9.86E-12 | *S100A8* | Protein coding | 24 |
| ENSG00000147570 | 2.12 | 2.82E-11 | *DNAJC5B* | Protein coding | 24 |
| ENSG00000100368 | 2.12 | 3.30E-31 | *CSF2RB* | Protein coding | 24 |
| ENSG00000087116 | 2.12 | 0.00151 | *ADAMTS2* | Protein coding | 24 |
| ENSG00000187848 | 2.12 | 0.00343 | *P2RX2* | Protein coding | 24 |
| ENSG00000196141 | 2.12 | 7.59E-08 | *SPATS2L* | Protein coding | 24 |
| ENSG00000127838 | 2.11 | 1.60E-37 | *PNKD* | Protein coding | 24 |
| ENSG00000160326 | 2.11 | 1.13E-24 | *SLC2A6* | Protein coding | 24 |
| ENSG00000184988 | 2.11 | 4.68E-38 | *TMEM106A* | Protein coding | 24 |
| ENSG00000124216 | 2.11 | 1.81E-05 | *SNAI1* | Protein coding | 24 |
| ENSG00000134575 | 2.10 | 2.66E-23 | *ACP2* | Protein coding | 24 |
| ENSG00000135124 | 2.10 | 1.98E-34 | *P2RX4* | Protein coding | 24 |
| ENSG00000095209 | 2.10 | 1.86E-35 | *TMEM38B* | Protein coding | 24 |
| ENSG00000188215 | 2.10 | 1.09E-33 | *DCUN1D3* | Protein coding | 24 |
| ENSG00000119922 | 2.09 | 6.62E-06 | *IFIT2* | Protein coding | 24 |
| ENSG00000160883 | 2.09 | 1.04E-09 | *HK3* | Protein coding | 24 |
| ENSG00000173535 | -2.09 | 6.51E-10 | *TNFRSF10C* | Protein coding | 24 |
| ENSG00000090376 | 2.09 | 4.91E-40 | *IRAK3* | Protein coding | 24 |
| ENSG00000178695 | 2.08 | 9.51E-54 | *KCTD12* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000121933 | -2.08 | 7.51E-11 | *ADORA3* | Protein coding | 24 |
| ENSG00000165694 | 2.08 | 0.00826 | *FRMD7* | Protein coding | 24 |
| ENSG00000100342 | 2.08 | 5.53E-23 | *APOL1* | Protein coding | 24 |
| ENSG00000214872 | 2.08 | 9.41E-18 | *SMTNL1* | Protein coding | 24 |
| ENSG00000175489 | 2.08 | 1.49E-31 | *LRRC25* | Protein coding | 24 |
| ENSG00000113494 | 2.07 | 1.09E-11 | *PRLR* | Protein coding | 24 |
| ENSG00000185745 | 2.07 | 0.00064 | *IFIT1* | Protein coding | 24 |
| ENSG00000142089 | 2.07 | 0.00011 | *IFITM3* | Protein coding | 24 |
| ENSG00000185222 | 2.07 | 3.28E-18 | *WBP5* | Protein coding | 24 |
| ENSG00000163625 | 2.07 | 4.44E-70 | *WDFY3* | Protein coding | 24 |
| ENSG00000136867 | 2.07 | 3.20E-16 | *SLC31A2* | Protein coding | 24 |
| ENSG00000137441 | -2.06 | 1.94E-39 | *FGFBP2* | Protein coding | 24 |
| ENSG00000197405 | 2.06 | 3.72E-38 | *C5AR1* | Protein coding | 24 |
| ENSG00000222047 | 2.06 | 0.00116 | *C10orf55* | Protein coding | 24 |
| ENSG00000064201 | -2.06 | 4.40E-74 | *TSPAN32* | Protein coding | 24 |
| ENSG00000116962 | 2.06 | 6.23E-20 | *NID1* | Protein coding | 24 |
| ENSG00000140274 | 2.06 | 0.00118 | *DUOXA2* | Protein coding | 24 |
| ENSG00000118785 | 2.06 | 9.41E-09 | *SPP1* | Protein coding | 24 |
| ENSG00000135604 | 2.05 | 4.56E-28 | *STX11* | Protein coding | 24 |
| ENSG00000255221 | 2.05 | 0.00024 | *CARD17* | Protein coding | 24 |
| ENSG00000020577 | 2.05 | 2.30E-17 | *SAMD4A* | Protein coding | 24 |
| ENSG00000176177 | 2.04 | 9.99E-05 | *ENTHD1* | Protein coding | 24 |
| ENSG00000106258 | 2.04 | 6.79E-13 | *CYP3A5* | Protein coding | 24 |
| ENSG00000248905 | 2.04 | 1.09E-10 | *FMN1* | Protein coding | 24 |
| ENSG00000172867 | -2.03 | 0.00012 | *KRT2* | Protein coding | 24 |
| ENSG00000168899 | 2.03 | 2.43E-19 | *VAMP5* | Protein coding | 24 |
| ENSG00000168658 | 2.03 | 0.00912 | *VWA3B* | Protein coding | 24 |
| ENSG00000249242 | 2.03 | 0.00018 | *TMEM150C* | Protein coding | 24 |
| ENSG00000132003 | 2.02 | 2.08E-38 | *ZSWIM4* | Protein coding | 24 |
| ENSG00000130589 | 2.02 | 1.95E-06 | *HELZ2* | Protein coding | 24 |
| ENSG00000181381 | 2.02 | 1.49E-29 | *DDX60L* | Protein coding | 24 |
| ENSG00000205300 | 2.01 | 0.00839 | *RP11-352D3.2* | Protein coding | 24 |
| ENSG00000137628 | 2.01 | 8.57E-12 | *DDX60* | Protein coding | 24 |
| ENSG00000125347 | 2.01 | 5.90E-35 | *IRF1* | Protein coding | 24 |
| ENSG00000105501 | 2.01 | 2.49E-21 | *SIGLEC5* | Protein coding | 24 |
| ENSG00000164111 | 2.01 | 4.55E-38 | *ANXA5* | Protein coding | 24 |
| ENSG00000221955 | 2.01 | 1.99E-09 | *SLC12A8* | Protein coding | 24 |
| ENSG00000164530 | -2.01 | 1.68E-13 | *PI16* | Protein coding | 24 |
| ENSG00000005102 | 2.01 | 4.37E-08 | *MEOX1* | Protein coding | 24 |
| ENSG00000156414 | 2.00 | 1.48E-16 | *TDRD9* | Protein coding | 24 |
| ENSG00000125462 | 2.00 | 8.41E-11 | *C1orf61* | Protein coding | 24 |
| ENSG00000157557 | 2.00 | 2.69E-36 | *ETS2* | Protein coding | 24 |
| ENSG00000141505 | -2.00 | 4.08E-13 | *ASGR1* | Protein coding | 24 |
| ENSG00000180539 | -2.00 | 2.29E-16 | *C9orf139* | Protein coding | 24 |
| ENSG00000187559 | -2.00 | 0.01075 | *FOXD4L3* | Protein coding | 24 |
| ENSG00000168329 | -2.00 | 3.50E-33 | *CX3CR1* | Protein coding | 24 |
| ENSG00000162614 | 2.00 | 1.64E-11 | *NEXN* | Protein coding | 24 |
| ENSG00000153071 | 1.99 | 2.91E-24 | *DAB2* | Protein coding | 24 |
| ENSG00000140749 | 1.99 | 2.69E-23 | *IGSF6* | Protein coding | 24 |
| ENSG00000131095 | 1.99 | 0.00569 | *GFAP* | Protein coding | 24 |
| ENSG00000135838 | 1.99 | 2.28E-24 | *NPL* | Protein coding | 24 |
| ENSG00000135454 | 1.99 | 0.00026 | *B4GALNT1* | Protein coding | 24 |
| ENSG00000166926 | 1.99 | 0.01002 | *MS4A6E* | Protein coding | 24 |
| ENSG00000137767 | 1.99 | 3.84E-28 | *SQRDL* | Protein coding | 24 |
| ENSG00000187608 | 1.98 | 0.00026 | *ISG15* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000152952 | 1.98 | 0.01217 | *PLOD2* | Protein coding | 24 |
| ENSG00000143479 | 1.98 | 2.63E-16 | *DYRK3* | Protein coding | 24 |
| ENSG00000104921 | 1.98 | 1.12E-25 | *FCER2* | Protein coding | 24 |
| ENSG00000186642 | 1.98 | 3.36E-10 | *PDE2A* | Protein coding | 24 |
| ENSG00000154734 | -1.97 | 7.26E-08 | *ADAMTS1* | Protein coding | 24 |
| ENSG00000183662 | -1.97 | 0.00558 | *FAM19A1* | Protein coding | 24 |
| ENSG00000101916 | 1.97 | 2.18E-39 | *TLR8* | Protein coding | 24 |
| ENSG00000173083 | 1.97 | 4.64E-31 | *HPSE* | Protein coding | 24 |
| ENSG00000047457 | 1.96 | 0.00038 | *CP* | Protein coding | 24 |
| ENSG00000131435 | 1.96 | 0.00159 | *PDLIM4* | Protein coding | 24 |
| ENSG00000167034 | 1.96 | 2.53E-15 | *NKX3-1* | Protein coding | 24 |
| ENSG00000204577 | 1.96 | 6.25E-19 | *LILRB3* | Protein coding | 24 |
| ENSG00000167613 | 1.96 | 2.69E-25 | *LAIR1* | Protein coding | 24 |
| ENSG00000182541 | 1.96 | 7.79E-46 | *LIMK2* | Protein coding | 24 |
| ENSG00000089169 | -1.96 | 0.01375 | *RPH3A* | Protein coding | 24 |
| ENSG00000137094 | 1.95 | 3.01E-28 | *DNAJB5* | Protein coding | 24 |
| ENSG00000150687 | -1.95 | 1.83E-22 | *PRSS23* | Protein coding | 24 |
| ENSG00000101321 | 1.95 | 0.00193 | *XKR7* | Protein coding | 24 |
| ENSG00000183960 | -1.95 | 1.74E-10 | *KCNH8* | Protein coding | 24 |
| ENSG00000117525 | 1.95 | 1.10E-13 | *F3* | Protein coding | 24 |
| ENSG00000068079 | 1.95 | 4.10E-07 | *IFI35* | Protein coding | 24 |
| ENSG00000197721 | 1.95 | 0.00029 | *CR1L* | Protein coding | 24 |
| ENSG00000135503 | 1.95 | 8.13E-28 | *ACVR1B* | Protein coding | 24 |
| ENSG00000163840 | 1.95 | 2.76E-25 | *DTX3L* | Protein coding | 24 |
| ENSG00000028137 | 1.95 | 5.37E-37 | *TNFRSF1B* | Protein coding | 24 |
| ENSG00000137710 | 1.95 | 3.95E-61 | *RDX* | Protein coding | 24 |
| ENSG00000178175 | 1.94 | 1.15E-12 | *ZNF366* | Protein coding | 24 |
| ENSG00000170458 | 1.94 | 8.09E-26 | *CD14* | Protein coding | 24 |
| ENSG00000111817 | 1.94 | 6.75E-39 | *DSE* | Protein coding | 24 |
| ENSG00000197702 | -1.94 | 0.00066 | *PARVA* | Protein coding | 24 |
| ENSG00000180875 | 1.94 | 3.66E-20 | *GREM2* | Protein coding | 24 |
| ENSG00000162772 | 1.94 | 5.22E-07 | *ATF3* | Protein coding | 24 |
| ENSG00000205358 | 1.94 | 0.00934 | *MT1H* | Protein coding | 24 |
| ENSG00000120875 | 1.93 | 7.21E-28 | *DUSP4* | Protein coding | 24 |
| ENSG00000111319 | -1.93 | 1.64E-07 | *SCNN1A* | Protein coding | 24 |
| ENSG00000137491 | 1.93 | 8.33E-08 | *SLCO2B1* | Protein coding | 24 |
| ENSG00000101384 | 1.93 | 9.09E-28 | *JAG1* | Protein coding | 24 |
| ENSG00000111424 | 1.93 | 1.46E-21 | *VDR* | Protein coding | 24 |
| ENSG00000166002 | 1.93 | 8.23E-08 | *C11orf75* | Protein coding | 24 |
| ENSG00000089692 | 1.93 | 6.98E-20 | *LAG3* | Protein coding | 24 |
| ENSG00000187726 | -1.92 | 4.36E-08 | *DNAJB13* | Protein coding | 24 |
| ENSG00000172794 | -1.92 | 6.19E-55 | *RAB37* | Protein coding | 24 |
| ENSG00000129673 | 1.92 | 3.80E-14 | *AANAT* | Protein coding | 24 |
| ENSG00000122733 | 1.92 | 0.01016 | *KIAA1045* | Protein coding | 24 |
| ENSG00000086548 | -1.92 | 5.21E-06 | *CEACAM6* | Protein coding | 24 |
| ENSG00000104974 | 1.92 | 9.76E-28 | *LILRA1* | Protein coding | 24 |
| ENSG00000154262 | -1.92 | 5.34E-10 | *ABCA6* | Protein coding | 24 |
| ENSG00000142611 | 1.91 | 0.00124 | *PRDM16* | Protein coding | 24 |
| ENSG00000157601 | 1.91 | 0.00037 | *MX1* | Protein coding | 24 |
| ENSG00000100079 | -1.91 | 0.00015 | *LGALS2* | Protein coding | 24 |
| ENSG00000183621 | 1.91 | 3.90E-40 | *ZNF438* | Protein coding | 24 |
| ENSG00000179165 | 1.91 | 0.00256 | *PXT1* | Protein coding | 24 |
| ENSG00000154258 | -1.91 | 3.57E-07 | *ABCA9* | Protein coding | 24 |
| ENSG00000213949 | 1.91 | 1.20E-71 | *ITGA1* | Protein coding | 24 |
| ENSG00000167858 | 1.91 | 0.01356 | *TEKT1* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000169136 | 1.91 | 1.09E-16 | *ATF5* | Protein coding | 24 |
| ENSG00000087085 | 1.91 | 1.48E-09 | *ACHE* | Protein coding | 24 |
| ENSG00000138642 | 1.91 | 3.81E-08 | *HERC6* | Protein coding | 24 |
| ENSG00000056661 | 1.90 | 0.00163 | *PCGF2* | Protein coding | 24 |
| ENSG00000163251 | 1.90 | 3.20E-13 | *FZD5* | Protein coding | 24 |
| ENSG00000186049 | -1.90 | 2.82E-26 | *KRT73* | Protein coding | 24 |
| ENSG00000104320 | 1.90 | 1.01E-95 | *NBN* | Protein coding | 24 |
| ENSG00000204001 | -1.90 | 4.54E-06 | *LCN8* | Protein coding | 24 |
| ENSG00000089127 | 1.90 | 0.00012 | *OAS1* | Protein coding | 24 |
| ENSG00000145936 | 1.89 | 2.08E-11 | *KCNMB1* | Protein coding | 24 |
| ENSG00000171509 | 1.89 | 0.01761 | *RXFP1* | Protein coding | 24 |
| ENSG00000180767 | -1.89 | 0.00153 | *CHST13* | Protein coding | 24 |
| ENSG00000169884 | -1.89 | 1.23E-10 | *WNT10B* | Protein coding | 24 |
| ENSG00000070729 | 1.89 | 0.01071 | *CNGB1* | Protein coding | 24 |
| ENSG00000217825 | 1.89 | 0.00942 | *AC099552.4* | Protein coding | 24 |
| ENSG00000135404 | 1.89 | 3.18E-22 | *CD63* | Protein coding | 24 |
| ENSG00000179148 | -1.89 | 0.01500 | *ALOXE3* | Protein coding | 24 |
| ENSG00000010327 | -1.89 | 4.37E-25 | *STAB1* | Protein coding | 24 |
| ENSG00000065328 | 1.89 | 0.00175 | *MCM10* | Protein coding | 24 |
| ENSG00000175471 | 1.88 | 6.38E-49 | *MCTP1* | Protein coding | 24 |
| ENSG00000060558 | 1.88 | 1.50E-21 | *GNA15* | Protein coding | 24 |
| ENSG00000197121 | 1.88 | 3.67E-40 | *PGAP1* | Protein coding | 24 |
| ENSG00000136830 | 1.87 | 5.19E-23 | *FAM129B* | Protein coding | 24 |
| ENSG00000105472 | -1.87 | 3.67E-07 | *CLEC11A* | Protein coding | 24 |
| ENSG00000145555 | 1.87 | 9.82E-08 | *MYO10* | Protein coding | 24 |
| ENSG00000136286 | 1.87 | 3.65E-32 | *MYO1G* | Protein coding | 24 |
| ENSG00000185885 | 1.87 | 3.59E-10 | *IFITM1* | Protein coding | 24 |
| ENSG00000230062 | 1.87 | 0.00986 | *ANKRD66* | Protein coding | 24 |
| ENSG00000112715 | 1.87 | 2.28E-07 | *VEGFA* | Protein coding | 24 |
| ENSG00000087074 | 1.87 | 5.86E-35 | *PPP1R15A* | Protein coding | 24 |
| ENSG00000169116 | -1.86 | 4.86E-15 | *PARM1* | Protein coding | 24 |
| ENSG00000122176 | -1.86 | 0.00113 | *FMOD* | Protein coding | 24 |
| ENSG00000143162 | 1.86 | 1.65E-26 | *CREG1* | Protein coding | 24 |
| ENSG00000074964 | 1.86 | 2.06E-09 | *ARHGEF10L* | Protein coding | 24 |
| ENSG00000156574 | -1.86 | 0.01152 | *NODAL* | Protein coding | 24 |
| ENSG00000155090 | 1.86 | 2.68E-53 | *KLF10* | Protein coding | 24 |
| ENSG00000205517 | -1.86 | 0.00295 | *RGL3* | Protein coding | 24 |
| ENSG00000107447 | -1.86 | 0.01159 | *DNTT* | Protein coding | 24 |
| ENSG00000143847 | -1.86 | 1.24E-13 | *PPFIA4* | Protein coding | 24 |
| ENSG00000160791 | 1.85 | 1.35E-20 | *CCR5* | Protein coding | 24 |
| ENSG00000137193 | 1.85 | 2.09E-53 | *PIM1* | Protein coding | 24 |
| ENSG00000162631 | -1.85 | 0.00530 | *NTNG1* | Protein coding | 24 |
| ENSG00000121690 | -1.85 | 4.27E-09 | *DEPDC7* | Protein coding | 24 |
| ENSG00000160213 | 1.85 | 4.89E-32 | *CSTB* | Protein coding | 24 |
| ENSG00000076716 | -1.85 | 2.61E-18 | *GPC4* | Protein coding | 24 |
| ENSG00000120949 | 1.85 | 2.79E-29 | *TNFRSF8* | Protein coding | 24 |
| ENSG00000179399 | 1.85 | 0.00183 | *GPC5* | Protein coding | 24 |
| ENSG00000105609 | 1.85 | 4.18E-08 | *LILRB5* | Protein coding | 24 |
| ENSG00000186529 | -1.85 | 4.03E-08 | *CYP4F3* | Protein coding | 24 |
| ENSG00000162383 | -1.85 | 9.58E-12 | *SLC1A7* | Protein coding | 24 |
| ENSG00000179750 | 1.85 | 4.50E-06 | *APOBEC3B* | Protein coding | 24 |
| ENSG00000163874 | 1.85 | 1.33E-46 | *ZC3H12A* | Protein coding | 24 |
| ENSG00000171621 | 1.85 | 1.37E-19 | *SPSB1* | Protein coding | 24 |
| ENSG00000135114 | 1.84 | 7.74E-06 | *OASL* | Protein coding | 24 |
| ENSG00000169495 | 1.84 | 0.00017 | *HTRA4* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000203685 | -1.84 | 1.26E-05 | *C1orf95* | Protein coding | 24 |
| ENSG00000087245 | 1.84 | 0.00054 | *MMP2* | Protein coding | 24 |
| ENSG00000185291 | 1.84 | 4.51E-20 | *IL3RA* | Protein coding | 24 |
| ENSG00000166922 | 1.83 | 1.89E-06 | *SCG5* | Protein coding | 24 |
| ENSG00000019991 | -1.83 | 2.18E-09 | *HGF* | Protein coding | 24 |
| ENSG00000102934 | -1.83 | 8.81E-09 | *PLLP* | Protein coding | 24 |
| ENSG00000166592 | 1.83 | 5.56E-07 | *RRAD* | Protein coding | 24 |
| ENSG00000173114 | -1.83 | 9.90E-19 | *LRRN3* | Protein coding | 24 |
| ENSG00000115267 | 1.83 | 4.89E-19 | *IFIH1* | Protein coding | 24 |
| ENSG00000087842 | 1.82 | 8.33E-10 | *PIR* | Protein coding | 24 |
| ENSG00000174460 | -1.82 | 0.01807 | *ZCCHC12* | Protein coding | 24 |
| ENSG00000166523 | 1.82 | 4.29E-23 | *CLEC4E* | Protein coding | 24 |
| ENSG00000011201 | 1.82 | 1.01E-05 | *KAL1* | Protein coding | 24 |
| ENSG00000115884 | 1.82 | 0.00360 | *SDC1* | Protein coding | 24 |
| ENSG00000173193 | 1.82 | 2.24E-24 | *PARP14* | Protein coding | 24 |
| ENSG00000179593 | 1.82 | 1.63E-15 | *ALOX15B* | Protein coding | 24 |
| ENSG00000173599 | -1.82 | 1.57E-38 | *PC* | Protein coding | 24 |
| ENSG00000010610 | -1.82 | 1.07E-45 | *CD4* | Protein coding | 24 |
| ENSG00000158481 | -1.82 | 3.25E-19 | *CD1C* | Protein coding | 24 |
| ENSG00000064932 | 1.82 | 1.10E-23 | *SBNO2* | Protein coding | 24 |
| ENSG00000203804 | 1.82 | 0.00014 | *C1orf138* | Protein coding | 24 |
| ENSG00000188158 | -1.82 | 1.88E-08 | *NHS* | Protein coding | 24 |
| ENSG00000142910 | -1.82 | 0.00344 | *TINAGL1* | Protein coding | 24 |
| ENSG00000141469 | -1.82 | 7.71E-18 | *SLC14A1* | Protein coding | 24 |
| ENSG00000178719 | 1.82 | 4.27E-25 | *GRINA* | Protein coding | 24 |
| ENSG00000163221 | 1.81 | 3.05E-15 | *S100A12* | Protein coding | 24 |
| ENSG00000180596 | 1.81 | 2.21E-09 | *HIST1H2BC* | Protein coding | 24 |
| ENSG00000183696 | 1.81 | 5.46E-41 | *UPP1* | Protein coding | 24 |
| ENSG00000167483 | -1.81 | 5.61E-54 | *FAM129C* | Protein coding | 24 |
| ENSG00000141574 | 1.81 | 7.83E-05 | *SECTM1* | Protein coding | 24 |
| ENSG00000241399 | -1.80 | 2.05E-33 | *CD302* | Protein coding | 24 |
| ENSG00000170801 | 1.80 | 0.00150 | *HTRA3* | Protein coding | 24 |
| ENSG00000173212 | 1.80 | 0.00600 | *MAB21L3* | Protein coding | 24 |
| ENSG00000169194 | 1.80 | 0.00094 | *IL13* | Protein coding | 24 |
| ENSG00000125430 | 1.79 | 2.25E-22 | *HS3ST3B1* | Protein coding | 24 |
| ENSG00000154102 | -1.79 | 1.51E-24 | *C16orf74* | Protein coding | 24 |
| ENSG00000187513 | 1.79 | 0.00665 | *GJA4* | Protein coding | 24 |
| ENSG00000186187 | 1.79 | 2.07E-32 | *ZNRF1* | Protein coding | 24 |
| ENSG00000158315 | 1.79 | 0.00754 | *RHBDL2* | Protein coding | 24 |
| ENSG00000007516 | -1.79 | 8.43E-36 | *BAIAP3* | Protein coding | 24 |
| ENSG00000140932 | -1.78 | 8.51E-05 | *CMTM2* | Protein coding | 24 |
| ENSG00000267174 | -1.78 | 0.00232 | *CTC-510F12.4* | Protein coding | 24 |
| ENSG00000133101 | 1.78 | 8.62E-08 | *CCNA1* | Protein coding | 24 |
| ENSG00000167105 | 1.78 | 0.00082 | *TMEM92* | Protein coding | 24 |
| ENSG00000172232 | -1.78 | 0.00111 | *AZU1* | Protein coding | 24 |
| ENSG00000140284 | 1.78 | 5.98E-08 | *SLC27A2* | Protein coding | 24 |
| ENSG00000102393 | 1.78 | 3.15E-27 | *GLA* | Protein coding | 24 |
| ENSG00000163082 | 1.78 | 1.91E-22 | *SGPP2* | Protein coding | 24 |
| ENSG00000136694 | -1.78 | 0.02312 | *IL36A* | Protein coding | 24 |
| ENSG00000148737 | 1.78 | 1.22E-20 | *TCF7L2* | Protein coding | 24 |
| ENSG00000144597 | 1.78 | 3.29E-42 | *EAF1* | Protein coding | 24 |
| ENSG00000165507 | 1.78 | 1.08E-15 | *C10orf10* | Protein coding | 24 |
| ENSG00000239998 | 1.78 | 9.08E-15 | *LILRA2* | Protein coding | 24 |
| ENSG00000134716 | 1.78 | 1.50E-05 | *CYP2J2* | Protein coding | 24 |
| ENSG00000167994 | 1.78 | 0.00036 | *RAB3IL1* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000143147 | 1.78 | 2.15E-10 | *GPR161* | Protein coding | 24 |
| ENSG00000069122 | 1.77 | 0.02148 | *GPR116* | Protein coding | 24 |
| ENSG00000130222 | 1.77 | 4.36E-05 | *GADD45G* | Protein coding | 24 |
| ENSG00000144802 | 1.77 | 1.91E-76 | *NFKBIZ* | Protein coding | 24 |
| ENSG00000120457 | -1.77 | 0.02925 | *KCNJ5* | Protein coding | 24 |
| ENSG00000090530 | -1.76 | 0.00649 | *LEPREL1* | Protein coding | 24 |
| ENSG00000175779 | -1.76 | 0.02195 | *C15orf53* | Protein coding | 24 |
| ENSG00000085514 | 1.76 | 5.22E-28 | *PILRA* | Protein coding | 24 |
| ENSG00000115718 | -1.76 | 1.16E-05 | *PROC* | Protein coding | 24 |
| ENSG00000130643 | -1.76 | 0.00232 | *CALY* | Protein coding | 24 |
| ENSG00000164647 | 1.76 | 0.00720 | *STEAP1* | Protein coding | 24 |
| ENSG00000082996 | 1.76 | 6.64E-50 | *RNF13* | Protein coding | 24 |
| ENSG00000160471 | 1.76 | 0.00176 | *COX6B2* | Protein coding | 24 |
| ENSG00000171236 | 1.76 | 9.82E-11 | *LRG1* | Protein coding | 24 |
| ENSG00000203814 | 1.76 | 0.00054 | *HIST2H2BF* | Protein coding | 24 |
| ENSG00000196581 | -1.76 | 0.00296 | *AJAP1* | Protein coding | 24 |
| ENSG00000005381 | -1.76 | 6.19E-10 | *MPO* | Protein coding | 24 |
| ENSG00000117984 | 1.75 | 1.52E-07 | *CTSD* | Protein coding | 24 |
| ENSG00000167261 | -1.75 | 7.23E-47 | *DPEP2* | Protein coding | 24 |
| ENSG00000138646 | 1.75 | 3.10E-05 | *HERC5* | Protein coding | 24 |
| ENSG00000039560 | 1.75 | 1.14E-09 | *RAI14* | Protein coding | 24 |
| ENSG00000115828 | 1.75 | 1.48E-19 | *QPCT* | Protein coding | 24 |
| ENSG00000101230 | -1.75 | 1.10E-06 | *ISM1* | Protein coding | 24 |
| ENSG00000113361 | -1.75 | 0.02836 | *CDH6* | Protein coding | 24 |
| ENSG00000197982 | 1.75 | 2.80E-23 | *C1orf122* | Protein coding | 24 |
| ENSG00000085117 | 1.75 | 7.54E-25 | *CD82* | Protein coding | 24 |
| ENSG00000106025 | -1.75 | 0.03173 | *TSPAN12* | Protein coding | 24 |
| ENSG00000075035 | -1.74 | 0.01342 | *WSCD2* | Protein coding | 24 |
| ENSG00000134042 | -1.74 | 0.03111 | *MRO* | Protein coding | 24 |
| ENSG00000156587 | 1.74 | 4.60E-09 | *UBE2L6* | Protein coding | 24 |
| ENSG00000213937 | -1.74 | 0.00474 | *CLDN9* | Protein coding | 24 |
| ENSG00000100644 | 1.74 | 6.46E-43 | *HIF1A* | Protein coding | 24 |
| ENSG00000185507 | 1.74 | 2.74E-05 | *IRF7* | Protein coding | 24 |
| ENSG00000135077 | 1.74 | 9.18E-23 | *HAVCR2* | Protein coding | 24 |
| ENSG00000108179 | 1.74 | 3.98E-25 | *PPIF* | Protein coding | 24 |
| ENSG00000090382 | -1.73 | 5.16E-27 | *LYZ* | Protein coding | 24 |
| ENSG00000168903 | -1.73 | 0.02509 | *BTNL3* | Protein coding | 24 |
| ENSG00000062716 | 1.73 | 8.39E-52 | *VMP1* | Protein coding | 24 |
| ENSG00000198053 | 1.73 | 1.36E-23 | *SIRPA* | Protein coding | 24 |
| ENSG00000137267 | 1.73 | 3.10E-09 | *TUBB2A* | Protein coding | 24 |
| ENSG00000216490 | 1.73 | 8.97E-16 | *IFI30* | Protein coding | 24 |
| ENSG00000152778 | 1.73 | 3.48E-18 | *IFIT5* | Protein coding | 24 |
| ENSG00000171517 | 1.73 | 0.00032 | *LPAR3* | Protein coding | 24 |
| ENSG00000065911 | 1.73 | 2.19E-33 | *MTHFD2* | Protein coding | 24 |
| ENSG00000115956 | 1.72 | 2.99E-24 | *PLEK* | Protein coding | 24 |
| ENSG00000162692 | 1.72 | 8.60E-08 | *VCAM1* | Protein coding | 24 |
| ENSG00000108582 | 1.72 | 1.03E-23 | *CPD* | Protein coding | 24 |
| ENSG00000182379 | 1.72 | 4.52E-05 | *NXPH4* | Protein coding | 24 |
| ENSG00000101425 | -1.72 | 6.82E-15 | *BPI* | Protein coding | 24 |
| ENSG00000168754 | -1.72 | 0.00070 | *FAM178B* | Protein coding | 24 |
| ENSG00000170099 | 1.72 | 0.03302 | *SERPINA6* | Protein coding | 24 |
| ENSG00000168386 | -1.72 | 2.21E-10 | *FILIP1L* | Protein coding | 24 |
| ENSG00000136040 | 1.72 | 3.33E-32 | *PLXNC1* | Protein coding | 24 |
| ENSG00000160200 | 1.72 | 1.34E-07 | *CBS* | Protein coding | 24 |
| ENSG00000158186 | 1.72 | 1.19E-21 | *MRAS* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000127507 | 1.72 | 4.51E-34 | *EMR2* | Protein coding | 24 |
| ENSG00000115159 | 1.72 | 7.51E-45 | *GPD2* | Protein coding | 24 |
| ENSG00000133067 | -1.71 | 5.39E-20 | *LGR6* | Protein coding | 24 |
| ENSG00000169442 | -1.71 | 1.03E-64 | *CD52* | Protein coding | 24 |
| ENSG00000186047 | 1.71 | 7.61E-09 | *DLEU7* | Protein coding | 24 |
| ENSG00000074181 | 1.71 | 3.17E-07 | *NOTCH3* | Protein coding | 24 |
| ENSG00000141748 | -1.70 | 0.02025 | *ARL5C* | Protein coding | 24 |
| ENSG00000168394 | 1.70 | 2.55E-19 | *TAP1* | Protein coding | 24 |
| ENSG00000126709 | 1.70 | 0.00515 | *IFI6* | Protein coding | 24 |
| ENSG00000170439 | 1.70 | 6.38E-05 | *METTL7B* | Protein coding | 24 |
| ENSG00000176597 | 1.70 | 3.00E-24 | *B3GNT5* | Protein coding | 24 |
| ENSG00000198910 | 1.70 | 3.22E-19 | *L1CAM* | Protein coding | 24 |
| ENSG00000136231 | 1.70 | 0.00020 | *IGF2BP3* | Protein coding | 24 |
| ENSG00000163746 | 1.70 | 0.00042 | *PLSCR2* | Protein coding | 24 |
| ENSG00000214193 | 1.70 | 4.71E-14 | *SH3D21* | Protein coding | 24 |
| ENSG00000186407 | 1.70 | 7.33E-24 | *CD300E* | Protein coding | 24 |
| ENSG00000196542 | -1.69 | 0.00781 | *SPTSSB* | Protein coding | 24 |
| ENSG00000156510 | -1.69 | 4.54E-06 | *HKDC1* | Protein coding | 24 |
| ENSG00000147416 | 1.69 | 1.70E-25 | *ATP6V1B2* | Protein coding | 24 |
| ENSG00000130962 | 1.69 | 0.00026 | *PRRG1* | Protein coding | 24 |
| ENSG00000152315 | 1.69 | 3.32E-07 | *KCNK13* | Protein coding | 24 |
| ENSG00000187479 | 1.69 | 0.03083 | *C11orf96* | Protein coding | 24 |
| ENSG00000163393 | 1.69 | 6.95E-22 | *SLC22A15* | Protein coding | 24 |
| ENSG00000173281 | 1.69 | 1.09E-48 | *PPP1R3B* | Protein coding | 24 |
| ENSG00000174469 | -1.69 | 3.35E-12 | *CNTNAP2* | Protein coding | 24 |
| ENSG00000120162 | 1.69 | 1.94E-15 | *MOB3B* | Protein coding | 24 |
| ENSG00000177807 | 1.69 | 1.50E-05 | *KCNJ10* | Protein coding | 24 |
| ENSG00000169122 | 1.68 | 1.46E-06 | *FAM110B* | Protein coding | 24 |
| ENSG00000142549 | 1.68 | 0.00327 | *IGLON5* | Protein coding | 24 |
| ENSG00000185947 | 1.68 | 1.72E-37 | *ZNF267* | Protein coding | 24 |
| ENSG00000170385 | 1.68 | 1.14E-60 | *SLC30A1* | Protein coding | 24 |
| ENSG00000156011 | 1.68 | 3.34E-33 | *PSD3* | Protein coding | 24 |
| ENSG00000198768 | 1.68 | 0.03322 | *APCDD1L* | Protein coding | 24 |
| ENSG00000153823 | 1.68 | 1.68E-26 | *PID1* | Protein coding | 24 |
| ENSG00000010278 | -1.68 | 3.07E-27 | *CD9* | Protein coding | 24 |
| ENSG00000163958 | 1.68 | 0.00139 | *ZDHHC19* | Protein coding | 24 |
| ENSG00000164604 | 1.68 | 0.00787 | *GPR85* | Protein coding | 24 |
| ENSG00000183722 | 1.67 | 0.00060 | *LHFP* | Protein coding | 24 |
| ENSG00000110665 | -1.67 | 1.32E-55 | *C11orf21* | Protein coding | 24 |
| ENSG00000141837 | 1.67 | 5.93E-10 | *CACNA1A* | Protein coding | 24 |
| ENSG00000026103 | 1.67 | 2.39E-25 | *FAS* | Protein coding | 24 |
| ENSG00000167995 | 1.67 | 1.69E-21 | *BEST1* | Protein coding | 24 |
| ENSG00000188060 | 1.67 | 1.89E-06 | *RAB42* | Protein coding | 24 |
| ENSG00000167987 | 1.66 | 6.63E-19 | *VPS37C* | Protein coding | 24 |
| ENSG00000165383 | 1.66 | 0.04022 | *LRRC18* | Protein coding | 24 |
| ENSG00000107317 | -1.66 | 7.18E-08 | *PTGDS* | Protein coding | 24 |
| ENSG00000185339 | 1.66 | 0.00134 | *TCN2* | Protein coding | 24 |
| ENSG00000145362 | 1.66 | 0.00076 | *ANK2* | Protein coding | 24 |
| ENSG00000182326 | 1.66 | 7.32E-06 | *C1S* | Protein coding | 24 |
| ENSG00000050767 | -1.66 | 0.00045 | *COL23A1* | Protein coding | 24 |
| ENSG00000168209 | 1.66 | 1.93E-17 | *DDIT4* | Protein coding | 24 |
| ENSG00000145365 | 1.66 | 7.75E-29 | *TIFA* | Protein coding | 24 |
| ENSG00000174370 | -1.65 | 1.61E-07 | *C11orf45* | Protein coding | 24 |
| ENSG00000205846 | 1.65 | 1.14E-10 | *CLEC6A* | Protein coding | 24 |
| ENSG00000158062 | -1.65 | 1.71E-60 | *UBXN11* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000242498 | -1.65 | 5.39E-16 | *C15orf38* | Protein coding | 24 |
| ENSG00000141096 | -1.65 | 1.76E-13 | *DPEP3* | Protein coding | 24 |
| ENSG00000050438 | 1.65 | 5.06E-28 | *SLC4A8* | Protein coding | 24 |
| ENSG00000169855 | 1.65 | 1.67E-09 | *ROBO1* | Protein coding | 24 |
| ENSG00000104375 | 1.65 | 6.29E-34 | *STK3* | Protein coding | 24 |
| ENSG00000196154 | -1.65 | 2.50E-34 | *S100A4* | Protein coding | 24 |
| ENSG00000050628 | 1.65 | 0.00049 | *PTGER3* | Protein coding | 24 |
| ENSG00000110446 | 1.65 | 3.52E-16 | *SLC15A3* | Protein coding | 24 |
| ENSG00000076351 | -1.65 | 1.33E-11 | *SLC46A1* | Protein coding | 24 |
| ENSG00000171798 | 1.64 | 0.00228 | *KNDC1* | Protein coding | 24 |
| ENSG00000167434 | -1.64 | 0.04080 | *CA4* | Protein coding | 24 |
| ENSG00000114019 | 1.64 | 0.03745 | *AMOTL2* | Protein coding | 24 |
| ENSG00000148834 | 1.64 | 1.40E-24 | *GSTO1* | Protein coding | 24 |
| ENSG00000196754 | 1.64 | 0.01112 | *S100A2* | Protein coding | 24 |
| ENSG00000124469 | -1.64 | 1.89E-09 | *CEACAM8* | Protein coding | 24 |
| ENSG00000198853 | 1.64 | 5.98E-21 | *RUSC2* | Protein coding | 24 |
| ENSG00000114698 | 1.64 | 0.02646 | *PLSCR4* | Protein coding | 24 |
| ENSG00000123094 | 1.64 | 3.93E-07 | *RASSF8* | Protein coding | 24 |
| ENSG00000185046 | -1.64 | 0.01514 | *ANKS1B* | Protein coding | 24 |
| ENSG00000103942 | -1.64 | 1.19E-19 | *HOMER2* | Protein coding | 24 |
| ENSG00000079385 | 1.64 | 4.64E-09 | *CEACAM1* | Protein coding | 24 |
| ENSG00000171051 | 1.64 | 3.86E-31 | *FPR1* | Protein coding | 24 |
| ENSG00000101327 | 1.64 | 0.04402 | *PDYN* | Protein coding | 24 |
| ENSG00000198673 | 1.64 | 9.70E-12 | *FAM19A2* | Protein coding | 24 |
| ENSG00000172380 | 1.63 | 0.02730 | *GNG12* | Protein coding | 24 |
| ENSG00000090104 | 1.63 | 1.14E-40 | *RGS1* | Protein coding | 24 |
| ENSG00000049768 | 1.63 | 5.57E-21 | *FOXP3* | Protein coding | 24 |
| ENSG00000165695 | 1.63 | 0.00033 | *AK8* | Protein coding | 24 |
| ENSG00000179348 | 1.63 | 3.22E-12 | *GATA2* | Protein coding | 24 |
| ENSG00000118849 | 1.63 | 0.01795 | *RARRES1* | Protein coding | 24 |
| ENSG00000087086 | 1.63 | 5.29E-17 | *FTL* | Protein coding | 24 |
| ENSG00000110218 | 1.63 | 4.34E-43 | *PANX1* | Protein coding | 24 |
| ENSG00000125744 | 1.63 | 2.68E-12 | *RTN2* | Protein coding | 24 |
| ENSG00000100678 | 1.63 | 0.00401 | *SLC8A3* | Protein coding | 24 |
| ENSG00000166670 | 1.63 | 0.03889 | *MMP10* | Protein coding | 24 |
| ENSG00000181585 | -1.63 | 6.43E-05 | *TMIE* | Protein coding | 24 |
| ENSG00000255582 | -1.62 | 0.00868 | *OR10G2* | Protein coding | 24 |
| ENSG00000196189 | 1.62 | 3.73E-06 | *SEMA4A* | Protein coding | 24 |
| ENSG00000009694 | -1.62 | 3.49E-16 | *TENM1* | Protein coding | 24 |
| ENSG00000196355 | -1.62 | 1.50E-18 | *AC021860.1* | Protein coding | 24 |
| ENSG00000111335 | 1.62 | 6.71E-06 | *OAS2* | Protein coding | 24 |
| ENSG00000167094 | -1.62 | 7.84E-18 | *TTC16* | Protein coding | 24 |
| ENSG00000119686 | 1.62 | 1.09E-10 | *FLVCR2* | Protein coding | 24 |
| ENSG00000072682 | 1.62 | 4.03E-06 | *P4HA2* | Protein coding | 24 |
| ENSG00000156453 | -1.61 | 1.27E-09 | *PCDH1* | Protein coding | 24 |
| ENSG00000114948 | -1.61 | 5.44E-16 | *ADAM23* | Protein coding | 24 |
| ENSG00000227051 | -1.61 | 1.14E-24 | *C14orf132* | Protein coding | 24 |
| ENSG00000111863 | 1.61 | 8.10E-37 | *ADTRP* | Protein coding | 24 |
| ENSG00000167553 | 1.61 | 1.72E-12 | *TUBA1C* | Protein coding | 24 |
| ENSG00000120262 | 1.61 | 3.70E-08 | *CCDC170* | Protein coding | 24 |
| ENSG00000143412 | -1.61 | 1.66E-18 | *ANXA9* | Protein coding | 24 |
| ENSG00000138760 | 1.61 | 1.32E-37 | *SCARB2* | Protein coding | 24 |
| ENSG00000196622 | -1.61 | 0.00106 | *RIMBP3* | Protein coding | 24 |
| ENSG00000140511 | 1.61 | 4.11E-17 | *HAPLN3* | Protein coding | 24 |
| ENSG00000171595 | -1.61 | 2.64E-05 | *DNAI2* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000120051 | 1.61 | 0.00015 | *CCDC147* | Protein coding | 24 |
| ENSG00000091129 | -1.60 | 0.00066 | *NRCAM* | Protein coding | 24 |
| ENSG00000236320 | -1.60 | 8.22E-13 | *SLFN14* | Protein coding | 24 |
| ENSG00000197561 | -1.60 | 0.02666 | *ELANE* | Protein coding | 24 |
| ENSG00000121297 | 1.60 | 2.37E-13 | *TSHZ3* | Protein coding | 24 |
| ENSG00000055332 | 1.60 | 7.41E-08 | *EIF2AK2* | Protein coding | 24 |
| ENSG00000135148 | 1.60 | 5.47E-18 | *TRAFD1* | Protein coding | 24 |
| ENSG00000145901 | 1.60 | 2.44E-30 | *TNIP1* | Protein coding | 24 |
| ENSG00000144843 | 1.60 | 2.45E-24 | *ADPRH* | Protein coding | 24 |
| ENSG00000139354 | 1.60 | 3.31E-17 | *GAS2L3* | Protein coding | 24 |
| ENSG00000244486 | -1.60 | 6.14E-06 | *SCARF2* | Protein coding | 24 |
| ENSG00000196396 | 1.60 | 6.46E-63 | *PTPN1* | Protein coding | 24 |
| ENSG00000136630 | 1.60 | 2.38E-13 | *HLX* | Protein coding | 24 |
| ENSG00000107518 | -1.60 | 0.04426 | *ATRNL1* | Protein coding | 24 |
| ENSG00000138411 | 1.59 | 6.63E-05 | *HECW2* | Protein coding | 24 |
| ENSG00000139410 | 1.59 | 0.00030 | *SDSL* | Protein coding | 24 |
| ENSG00000118985 | 1.59 | 2.90E-38 | *ELL2* | Protein coding | 24 |
| ENSG00000101160 | 1.59 | 5.93E-16 | *CTSZ* | Protein coding | 24 |
| ENSG00000005108 | -1.59 | 0.00884 | *THSD7A* | Protein coding | 24 |
| ENSG00000187583 | 1.59 | 3.36E-12 | *PLEKHN1* | Protein coding | 24 |
| ENSG00000158104 | 1.59 | 0.00211 | *HPD* | Protein coding | 24 |
| ENSG00000198515 | -1.58 | 0.02740 | *CNGA1* | Protein coding | 24 |
| ENSG00000198785 | 1.58 | 6.78E-06 | *GRIN3A* | Protein coding | 24 |
| ENSG00000242616 | 1.58 | 2.21E-07 | *GNG10* | Protein coding | 24 |
| ENSG00000031081 | 1.58 | 1.14E-18 | *ARHGAP31* | Protein coding | 24 |
| ENSG00000080854 | -1.58 | 4.93E-16 | *IGSF9B* | Protein coding | 24 |
| ENSG00000167895 | -1.58 | 1.02E-43 | *TMC8* | Protein coding | 24 |
| ENSG00000197705 | -1.58 | 3.91E-20 | *KLHL14* | Protein coding | 24 |
| ENSG00000100097 | 1.58 | 5.38E-15 | *LGALS1* | Protein coding | 24 |
| ENSG00000176641 | 1.58 | 0.00610 | *RNF152* | Protein coding | 24 |
| ENSG00000130487 | 1.58 | 2.33E-10 | *KLHDC7B* | Protein coding | 24 |
| ENSG00000103257 | 1.58 | 1.02E-27 | *SLC7A5* | Protein coding | 24 |
| ENSG00000197461 | 1.58 | 4.56E-07 | *PDGFA* | Protein coding | 24 |
| ENSG00000185432 | -1.58 | 9.16E-27 | *METTL7A* | Protein coding | 24 |
| ENSG00000178226 | -1.58 | 1.48E-09 | *PRSS36* | Protein coding | 24 |
| ENSG00000073737 | -1.58 | 1.09E-06 | *DHRS9* | Protein coding | 24 |
| ENSG00000164713 | 1.57 | 1.08E-21 | *BRI3* | Protein coding | 24 |
| ENSG00000168461 | 1.57 | 3.48E-30 | *RAB31* | Protein coding | 24 |
| ENSG00000161640 | 1.57 | 0.00205 | *SIGLEC11* | Protein coding | 24 |
| ENSG00000053108 | -1.57 | 4.32E-05 | *FSTL4* | Protein coding | 24 |
| ENSG00000184060 | 1.57 | 6.28E-16 | *ADAP2* | Protein coding | 24 |
| ENSG00000170956 | 1.57 | 8.25E-10 | *CEACAM3* | Protein coding | 24 |
| ENSG00000115592 | 1.57 | 0.00074 | *PRKAG3* | Protein coding | 24 |
| ENSG00000179604 | 1.57 | 5.03E-08 | *CDC42EP4* | Protein coding | 24 |
| ENSG00000102265 | 1.57 | 1.82E-15 | *TIMP1* | Protein coding | 24 |
| ENSG00000166546 | -1.57 | 0.01330 | *BEAN1* | Protein coding | 24 |
| ENSG00000103642 | 1.57 | 1.82E-32 | *LACTB* | Protein coding | 24 |
| ENSG00000162645 | 1.56 | 1.37E-23 | *GBP2* | Protein coding | 24 |
| ENSG00000110076 | 1.56 | 8.80E-08 | *NRXN2* | Protein coding | 24 |
| ENSG00000102580 | 1.56 | 4.57E-30 | *DNAJC3* | Protein coding | 24 |
| ENSG00000168389 | 1.56 | 1.54E-24 | *MFSD2A* | Protein coding | 24 |
| ENSG00000159374 | -1.56 | 0.00873 | *M1AP* | Protein coding | 24 |
| ENSG00000139629 | 1.56 | 3.63E-23 | *GALNT6* | Protein coding | 24 |
| ENSG00000196132 | -1.56 | 0.00162 | *MYT1* | Protein coding | 24 |
| ENSG00000118160 | 1.56 | 0.03216 | *SLC8A2* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000196562 | -1.56 | 1.04E-14 | *SULF2* | Protein coding | 24 |
| ENSG00000069188 | -1.56 | 6.02E-13 | *SDK2* | Protein coding | 24 |
| ENSG00000164512 | -1.56 | 8.43E-09 | *ANKRD55* | Protein coding | 24 |
| ENSG00000183742 | 1.56 | 8.27E-11 | *MACC1* | Protein coding | 24 |
| ENSG00000148175 | 1.55 | 1.25E-29 | *STOM* | Protein coding | 24 |
| ENSG00000197262 | 1.55 | 0.00415 | *CCL4L2* | Protein coding | 24 |
| ENSG00000117215 | -1.55 | 0.04837 | *PLA2G2D* | Protein coding | 24 |
| ENSG00000071967 | -1.55 | 2.01E-23 | *CYBRD1* | Protein coding | 24 |
| ENSG00000153976 | 1.55 | 0.00981 | *HS3ST3A1* | Protein coding | 24 |
| ENSG00000214402 | -1.55 | 0.01775 | *LCNL1* | Protein coding | 24 |
| ENSG00000157470 | -1.55 | 0.00303 | *FAM81A* | Protein coding | 24 |
| ENSG00000104783 | 1.55 | 7.06E-32 | *KCNN4* | Protein coding | 24 |
| ENSG00000244242 | 1.55 | 2.25E-07 | *IFITM10* | Protein coding | 24 |
| ENSG00000166510 | -1.55 | 3.25E-07 | *CCDC68* | Protein coding | 24 |
| ENSG00000140939 | 1.55 | 6.52E-12 | *NOL3* | Protein coding | 24 |
| ENSG00000128335 | 1.54 | 2.13E-26 | *APOL2* | Protein coding | 24 |
| ENSG00000181649 | 1.54 | 0.00096 | *PHLDA2* | Protein coding | 24 |
| ENSG00000117226 | 1.54 | 7.59E-33 | *GBP3* | Protein coding | 24 |
| ENSG00000255346 | -1.54 | 0.04765 | *NOX5* | Protein coding | 24 |
| ENSG00000171435 | -1.54 | 0.03073 | *KSR2* | Protein coding | 24 |
| ENSG00000132109 | 1.54 | 1.45E-17 | *TRIM21* | Protein coding | 24 |
| ENSG00000172403 | 1.54 | 1.09E-13 | *SYNPO2* | Protein coding | 24 |
| ENSG00000160145 | -1.54 | 3.81E-08 | *KALRN* | Protein coding | 24 |
| ENSG00000118515 | 1.54 | 1.48E-28 | *SGK1* | Protein coding | 24 |
| ENSG00000227184 | -1.53 | 1.60E-15 | *EPPK1* | Protein coding | 24 |
| ENSG00000140564 | 1.53 | 7.39E-25 | *FURIN* | Protein coding | 24 |
| ENSG00000163354 | -1.53 | 4.11E-06 | *DCST2* | Protein coding | 24 |
| ENSG00000148459 | 1.53 | 1.63E-14 | *PDSS1* | Protein coding | 24 |
| ENSG00000166016 | 1.53 | 1.62E-17 | *ABTB2* | Protein coding | 24 |
| ENSG00000087237 | 1.53 | 2.94E-10 | *CETP* | Protein coding | 24 |
| ENSG00000081985 | 1.53 | 4.72E-30 | *IL12RB2* | Protein coding | 24 |
| ENSG00000141655 | -1.53 | 4.64E-17 | *TNFRSF11A* | Protein coding | 24 |
| ENSG00000169439 | 1.53 | 2.70E-23 | *SDC2* | Protein coding | 24 |
| ENSG00000174326 | -1.53 | 1.22E-05 | *SLC16A11* | Protein coding | 24 |
| ENSG00000172070 | 1.53 | 3.15E-17 | *SRXN1* | Protein coding | 24 |
| ENSG00000110799 | -1.53 | 9.62E-05 | *VWF* | Protein coding | 24 |
| ENSG00000109756 | 1.53 | 1.34E-60 | *RAPGEF2* | Protein coding | 24 |
| ENSG00000099860 | 1.52 | 1.90E-21 | *GADD45B* | Protein coding | 24 |
| ENSG00000114812 | -1.52 | 5.76E-44 | *VIPR1* | Protein coding | 24 |
| ENSG00000124743 | -1.52 | 0.00897 | *KLHL31* | Protein coding | 24 |
| ENSG00000110203 | -1.52 | 5.15E-06 | *FOLR3* | Protein coding | 24 |
| ENSG00000123130 | 1.52 | 1.61E-25 | *ACOT9* | Protein coding | 24 |
| ENSG00000111261 | -1.52 | 0.00814 | *MANSC1* | Protein coding | 24 |
| ENSG00000182022 | 1.52 | 1.08E-28 | *CHST15* | Protein coding | 24 |
| ENSG00000070614 | 1.52 | 9.38E-21 | *NDST1* | Protein coding | 24 |
| ENSG00000086288 | -1.52 | 1.40E-10 | *NME8* | Protein coding | 24 |
| ENSG00000188157 | 1.52 | 2.59E-05 | *AGRN* | Protein coding | 24 |
| ENSG00000166927 | 1.52 | 1.74E-23 | *MS4A7* | Protein coding | 24 |
| ENSG00000132386 | -1.52 | 1.55E-23 | *SERPINF1* | Protein coding | 24 |
| ENSG00000167711 | -1.52 | 1.02E-09 | *SERPINF2* | Protein coding | 24 |
| ENSG00000089012 | -1.52 | 2.71E-21 | *SIRPG* | Protein coding | 24 |
| ENSG00000164251 | 1.52 | 1.85E-13 | *F2RL1* | Protein coding | 24 |
| ENSG00000186522 | 1.52 | 3.72E-18 | *Sep-10* | Protein coding | 24 |
| ENSG00000129667 | 1.52 | 2.04E-17 | *RHBDF2* | Protein coding | 24 |
| ENSG00000197723 | 1.51 | 0.02724 | *HSPB9* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000113368 | 1.51 | 4.55E-21 | *LMNB1* | Protein coding | 24 |
| ENSG00000170581 | 1.51 | 1.42E-14 | *STAT2* | Protein coding | 24 |
| ENSG00000126003 | 1.51 | 5.51E-51 | *PLAGL2* | Protein coding | 24 |
| ENSG00000107201 | 1.51 | 1.94E-16 | *DDX58* | Protein coding | 24 |
| ENSG00000196167 | -1.51 | 0.00866 | *C11orf92* | Protein coding | 24 |
| ENSG00000092051 | 1.51 | 0.04393 | *JPH4* | Protein coding | 24 |
| ENSG00000151364 | 1.51 | 0.03464 | *KCTD14* | Protein coding | 24 |
| ENSG00000132274 | 1.51 | 1.77E-15 | *TRIM22* | Protein coding | 24 |
| ENSG00000186204 | -1.51 | 0.01338 | *CYP4F12* | Protein coding | 24 |
| ENSG00000156273 | 1.51 | 1.58E-57 | *BACH1* | Protein coding | 24 |
| ENSG00000140470 | 1.50 | 6.41E-08 | *ADAMTS17* | Protein coding | 24 |
| ENSG00000169604 | -1.50 | 0.01882 | *ANTXR1* | Protein coding | 24 |
| ENSG00000179294 | 1.50 | 2.28E-09 | *C17orf96* | Protein coding | 24 |
| ENSG00000119714 | 1.50 | 5.91E-14 | *GPR68* | Protein coding | 24 |
| ENSG00000108679 | 1.50 | 1.05E-06 | *LGALS3BP* | Protein coding | 24 |
| ENSG00000110880 | 1.50 | 1.10E-23 | *CORO1C* | Protein coding | 24 |
| ENSG00000196743 | 1.50 | 2.47E-18 | *GM2A* | Protein coding | 24 |
| ENSG00000183486 | 1.49 | 0.00011 | *MX2* | Protein coding | 24 |
| ENSG00000007314 | -1.49 | 0.00042 | *SCN4A* | Protein coding | 24 |
| ENSG00000177398 | -1.49 | 0.01558 | *UMODL1* | Protein coding | 24 |
| ENSG00000136868 | 1.49 | 1.14E-23 | *SLC31A1* | Protein coding | 24 |
| ENSG00000184985 | 1.49 | 0.00145 | *SORCS2* | Protein coding | 24 |
| ENSG00000169435 | -1.49 | 1.67E-07 | *RASSF6* | Protein coding | 24 |
| ENSG00000184838 | 1.49 | 1.06E-05 | *PRR16* | Protein coding | 24 |
| ENSG00000114923 | -1.49 | 0.03207 | *SLC4A3* | Protein coding | 24 |
| ENSG00000156127 | 1.49 | 4.13E-15 | *BATF* | Protein coding | 24 |
| ENSG00000025708 | 1.48 | 0.00020 | *TYMP* | Protein coding | 24 |
| ENSG00000197712 | 1.48 | 5.59E-10 | *FAM114A1* | Protein coding | 24 |
| ENSG00000241978 | 1.48 | 0.00060 | *AKAP2* | Protein coding | 24 |
| ENSG00000187775 | 1.48 | 0.00015 | *DNAH17* | Protein coding | 24 |
| ENSG00000165272 | -1.48 | 9.49E-63 | *AQP3* | Protein coding | 24 |
| ENSG00000130881 | -1.48 | 0.00036 | *LRP3* | Protein coding | 24 |
| ENSG00000198467 | -1.48 | 4.64E-54 | *TPM2* | Protein coding | 24 |
| ENSG00000179057 | -1.48 | 0.00052 | *IGSF22* | Protein coding | 24 |
| ENSG00000213759 | -1.48 | 0.01603 | *UGT2B11* | Protein coding | 24 |
| ENSG00000121316 | -1.48 | 1.99E-11 | *PLBD1* | Protein coding | 24 |
| ENSG00000165125 | -1.48 | 0.00324 | *TRPV6* | Protein coding | 24 |
| ENSG00000100600 | 1.48 | 4.98E-10 | *LGMN* | Protein coding | 24 |
| ENSG00000143322 | 1.47 | 4.13E-55 | *ABL2* | Protein coding | 24 |
| ENSG00000170909 | 1.47 | 2.96E-06 | *OSCAR* | Protein coding | 24 |
| ENSG00000154269 | 1.47 | 0.00014 | *ENPP3* | Protein coding | 24 |
| ENSG00000070985 | -1.47 | 0.02786 | *TRPM5* | Protein coding | 24 |
| ENSG00000129450 | 1.47 | 2.33E-35 | *SIGLEC9* | Protein coding | 24 |
| ENSG00000188888 | -1.47 | 0.02661 | *GPR179* | Protein coding | 24 |
| ENSG00000120057 | -1.47 | 0.00265 | *SFRP5* | Protein coding | 24 |
| ENSG00000177453 | 1.47 | 0.00052 | *NIM1* | Protein coding | 24 |
| ENSG00000204978 | -1.47 | 0.03836 | *C19orf69* | Protein coding | 24 |
| ENSG00000142621 | 1.47 | 1.36E-06 | *FHAD1* | Protein coding | 24 |
| ENSG00000100346 | -1.47 | 1.37E-46 | *CACNA1I* | Protein coding | 24 |
| ENSG00000160233 | 1.47 | 4.24E-08 | *LRRC3* | Protein coding | 24 |
| ENSG00000072401 | 1.47 | 8.15E-27 | *UBE2D1* | Protein coding | 24 |
| ENSG00000102921 | 1.47 | 3.60E-43 | *N4BP1* | Protein coding | 24 |
| ENSG00000163191 | 1.46 | 6.59E-18 | *S100A11* | Protein coding | 24 |
| ENSG00000161509 | -1.46 | 0.00023 | *GRIN2C* | Protein coding | 24 |
| ENSG00000073331 | 1.46 | 8.06E-21 | *ALPK1* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000153317 | 1.46 | 7.32E-42 | *ASAP1* | Protein coding | 24 |
| ENSG00000166963 | 1.46 | 3.39E-09 | *MAP1A* | Protein coding | 24 |
| ENSG00000117115 | -1.46 | 2.37E-11 | *PADI2* | Protein coding | 24 |
| ENSG00000135929 | 1.46 | 5.47E-13 | *CYP27A1* | Protein coding | 24 |
| ENSG00000164823 | 1.46 | 1.16E-27 | *OSGIN2* | Protein coding | 24 |
| ENSG00000165181 | 1.46 | 0.01289 | *C9orf84* | Protein coding | 24 |
| ENSG00000198715 | 1.46 | 4.52E-13 | *C1orf85* | Protein coding | 24 |
| ENSG00000171094 | 1.46 | 0.00844 | *ALK* | Protein coding | 24 |
| ENSG00000162723 | 1.46 | 0.00046 | *SLAMF9* | Protein coding | 24 |
| ENSG00000179241 | 1.45 | 2.51E-13 | *LDLRAD3* | Protein coding | 24 |
| ENSG00000064886 | 1.45 | 5.61E-15 | *CHI3L2* | Protein coding | 24 |
| ENSG00000254918 | -1.45 | 1.72E-15 | *RP11-259P6.1* | Protein coding | 24 |
| ENSG00000131370 | 1.45 | 8.49E-19 | *SH3BP5* | Protein coding | 24 |
| ENSG00000073150 | 1.45 | 1.11E-07 | *PANX2* | Protein coding | 24 |
| ENSG00000185386 | 1.45 | 2.13E-14 | *MAPK11* | Protein coding | 24 |
| ENSG00000168003 | 1.45 | 1.95E-18 | *SLC3A2* | Protein coding | 24 |
| ENSG00000146021 | -1.45 | 5.54E-19 | *KLHL3* | Protein coding | 24 |
| ENSG00000142657 | 1.45 | 3.82E-17 | *PGD* | Protein coding | 24 |
| ENSG00000047249 | 1.44 | 1.97E-24 | *ATP6V1H* | Protein coding | 24 |
| ENSG00000157551 | 1.44 | 1.29E-17 | *KCNJ15* | Protein coding | 24 |
| ENSG00000154764 | -1.44 | 1.81E-16 | *WNT7A* | Protein coding | 24 |
| ENSG00000171631 | 1.44 | 0.00032 | *P2RY6* | Protein coding | 24 |
| ENSG00000151948 | 1.44 | 7.24E-13 | *GLT1D1* | Protein coding | 24 |
| ENSG00000101017 | 1.44 | 1.03E-17 | *CD40* | Protein coding | 24 |
| ENSG00000169504 | 1.44 | 1.13E-31 | *CLIC4* | Protein coding | 24 |
| ENSG00000089351 | 1.44 | 2.42E-20 | *GRAMD1A* | Protein coding | 24 |
| ENSG00000133055 | 1.44 | 4.03E-06 | *MYBPH* | Protein coding | 24 |
| ENSG00000162398 | -1.43 | 0.02769 | *C1orf177* | Protein coding | 24 |
| ENSG00000177628 | 1.43 | 5.01E-18 | *GBA* | Protein coding | 24 |
| ENSG00000170667 | -1.43 | 2.04E-05 | *RASA4B* | Protein coding | 24 |
| ENSG00000124785 | 1.43 | 0.00494 | *NRN1* | Protein coding | 24 |
| ENSG00000172548 | 1.43 | 0.00253 | *NIPAL4* | Protein coding | 24 |
| ENSG00000212734 | -1.43 | 0.02178 | *C17orf100* | Protein coding | 24 |
| ENSG00000066697 | 1.43 | 7.65E-20 | *MSANTD3* | Protein coding | 24 |
| ENSG00000196116 | 1.43 | 3.40E-16 | *TDRD7* | Protein coding | 24 |
| ENSG00000103647 | -1.43 | 0.00105 | *CORO2B* | Protein coding | 24 |
| ENSG00000088827 | 1.43 | 0.03623 | *SIGLEC1* | Protein coding | 24 |
| ENSG00000197081 | 1.42 | 1.60E-23 | *IGF2R* | Protein coding | 24 |
| ENSG00000143226 | 1.42 | 3.69E-24 | *FCGR2A* | Protein coding | 24 |
| ENSG00000183856 | -1.42 | 0.00029 | *IQGAP3* | Protein coding | 24 |
| ENSG00000114450 | 1.42 | 9.24E-29 | *GNB4* | Protein coding | 24 |
| ENSG00000164136 | 1.42 | 1.68E-17 | *IL15* | Protein coding | 24 |
| ENSG00000143153 | 1.42 | 1.50E-06 | *ATP1B1* | Protein coding | 24 |
| ENSG00000089250 | -1.42 | 0.04434 | *NOS1* | Protein coding | 24 |
| ENSG00000130948 | -1.42 | 0.00717 | *HSD17B3* | Protein coding | 24 |
| ENSG00000163568 | 1.42 | 1.17E-18 | *AIM2* | Protein coding | 24 |
| ENSG00000259030 | -1.42 | 0.00222 | *FPGT-TNNI3K* | Protein coding | 24 |
| ENSG00000069399 | 1.42 | 3.32E-18 | *BCL3* | Protein coding | 24 |
| ENSG00000104763 | 1.42 | 1.16E-20 | *ASAH1* | Protein coding | 24 |
| ENSG00000138600 | 1.42 | 1.82E-53 | *SPPL2A* | Protein coding | 24 |
| ENSG00000183473 | 1.42 | 8.23E-14 | *SSTR3* | Protein coding | 24 |
| ENSG00000085871 | -1.42 | 0.00011 | *MGST2* | Protein coding | 24 |
| ENSG00000114127 | 1.42 | 2.63E-57 | *XRN1* | Protein coding | 24 |
| ENSG00000167281 | -1.41 | 0.00478 | *RBFOX3* | Protein coding | 24 |
| ENSG00000144118 | 1.41 | 1.05E-25 | *RALB* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000161647 | 1.41 | 0.00164 | *MPP3* | Protein coding | 24 |
| ENSG00000151704 | -1.41 | 0.01160 | *KCNJ1* | Protein coding | 24 |
| ENSG00000187642 | 1.41 | 0.00027 | *C1orf170* | Protein coding | 24 |
| ENSG00000133805 | 1.41 | 1.50E-22 | *AMPD3* | Protein coding | 24 |
| ENSG00000136869 | 1.41 | 4.97E-24 | *TLR4* | Protein coding | 24 |
| ENSG00000143434 | -1.41 | 5.13E-15 | *SEMA6C* | Protein coding | 24 |
| ENSG00000170542 | 1.41 | 1.93E-33 | *SERPINB9* | Protein coding | 24 |
| ENSG00000152128 | 1.41 | 0.00015 | *TMEM163* | Protein coding | 24 |
| ENSG00000145506 | 1.41 | 0.03841 | *NKD2* | Protein coding | 24 |
| ENSG00000196730 | 1.41 | 3.08E-13 | *DAPK1* | Protein coding | 24 |
| ENSG00000026508 | 1.40 | 1.95E-33 | *CD44* | Protein coding | 24 |
| ENSG00000173221 | 1.40 | 4.34E-28 | *GLRX* | Protein coding | 24 |
| ENSG00000056972 | 1.40 | 4.52E-13 | *TRAF3IP2* | Protein coding | 24 |
| ENSG00000215244 | 1.40 | 8.42E-06 | *AL137145.1* | Protein coding | 24 |
| ENSG00000121764 | 1.40 | 0.01228 | *HCRTR1* | Protein coding | 24 |
| ENSG00000173821 | 1.40 | 3.29E-23 | *RNF213* | Protein coding | 24 |
| ENSG00000170959 | -1.40 | 0.03313 | *DCDC5* | Protein coding | 24 |
| ENSG00000102755 | 1.40 | 2.80E-07 | *FLT1* | Protein coding | 24 |
| ENSG00000160932 | 1.40 | 0.00315 | *LY6E* | Protein coding | 24 |
| ENSG00000196954 | 1.40 | 4.64E-54 | *CASP4* | Protein coding | 24 |
| ENSG00000169507 | -1.40 | 0.00032 | *SLC38A11* | Protein coding | 24 |
| ENSG00000020181 | 1.39 | 3.30E-08 | *GPR124* | Protein coding | 24 |
| ENSG00000106546 | 1.39 | 9.71E-52 | *AHR* | Protein coding | 24 |
| ENSG00000107968 | 1.39 | 4.14E-45 | *MAP3K8* | Protein coding | 24 |
| ENSG00000119655 | 1.39 | 3.77E-13 | *NPC2* | Protein coding | 24 |
| ENSG00000163235 | 1.39 | 9.27E-14 | *TGFA* | Protein coding | 24 |
| ENSG00000249992 | 1.39 | 0.00819 | *TMEM158* | Protein coding | 24 |
| ENSG00000205730 | 1.39 | 1.81E-17 | *ITPRIPL2* | Protein coding | 24 |
| ENSG00000135677 | 1.39 | 1.78E-30 | *GNS* | Protein coding | 24 |
| ENSG00000174600 | 1.39 | 9.15E-05 | *CMKLR1* | Protein coding | 24 |
| ENSG00000103811 | 1.39 | 1.10E-12 | *CTSH* | Protein coding | 24 |
| ENSG00000106976 | -1.38 | 0.00070 | *DNM1* | Protein coding | 24 |
| ENSG00000184731 | -1.38 | 5.43E-07 | *FAM110C* | Protein coding | 24 |
| ENSG00000183134 | -1.38 | 5.92E-06 | *PTGDR2* | Protein coding | 24 |
| ENSG00000132205 | 1.38 | 4.71E-20 | *EMILIN2* | Protein coding | 24 |
| ENSG00000186088 | 1.38 | 2.57E-26 | *PION* | Protein coding | 24 |
| ENSG00000108771 | 1.38 | 8.06E-05 | *DHX58* | Protein coding | 24 |
| ENSG00000170379 | 1.38 | 3.78E-19 | *FAM115C* | Protein coding | 24 |
| ENSG00000135116 | -1.38 | 4.85E-06 | *HRK* | Protein coding | 24 |
| ENSG00000197142 | 1.38 | 2.85E-64 | *ACSL5* | Protein coding | 24 |
| ENSG00000139610 | -1.38 | 0.02914 | *CELA1* | Protein coding | 24 |
| ENSG00000120896 | -1.38 | 7.93E-46 | *SORBS3* | Protein coding | 24 |
| ENSG00000137098 | -1.38 | 4.31E-06 | *SPAG8* | Protein coding | 24 |
| ENSG00000174939 | -1.37 | 0.03848 | *ASPHD1* | Protein coding | 24 |
| ENSG00000185909 | 1.37 | 6.02E-06 | *KLHDC8B* | Protein coding | 24 |
| ENSG00000258227 | 1.37 | 3.28E-05 | *CLEC5A* | Protein coding | 24 |
| ENSG00000102048 | -1.37 | 0.04733 | *ASB9* | Protein coding | 24 |
| ENSG00000186074 | 1.37 | 8.42E-05 | *CD300LF* | Protein coding | 24 |
| ENSG00000156869 | 1.37 | 0.00021 | *FRRS1* | Protein coding | 24 |
| ENSG00000110911 | 1.36 | 8.28E-46 | *SLC11A2* | Protein coding | 24 |
| ENSG00000185187 | -1.36 | 4.75E-36 | *SIGIRR* | Protein coding | 24 |
| ENSG00000183773 | -1.36 | 7.11E-05 | *AIFM3* | Protein coding | 24 |
| ENSG00000196517 | 1.36 | 4.58E-05 | *SLC6A9* | Protein coding | 24 |
| ENSG00000100911 | 1.36 | 8.96E-14 | *PSME2* | Protein coding | 24 |
| ENSG00000127947 | 1.36 | 2.98E-27 | *PTPN12* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000135519 | -1.36 | 4.00E-10 | *KCNH3* | Protein coding | 24 |
| ENSG00000267534 | 1.36 | 5.65E-08 | *S1PR2* | Protein coding | 24 |
| ENSG00000172426 | 1.36 | 0.00020 | *RSPH9* | Protein coding | 24 |
| ENSG00000111110 | 1.36 | 1.35E-06 | *PPM1H* | Protein coding | 24 |
| ENSG00000172183 | 1.36 | 6.54E-13 | *ISG20* | Protein coding | 24 |
| ENSG00000141458 | 1.36 | 1.16E-28 | *NPC1* | Protein coding | 24 |
| ENSG00000162482 | -1.36 | 0.01327 | *AKR7A3* | Protein coding | 24 |
| ENSG00000102554 | 1.36 | 2.72E-12 | *KLF5* | Protein coding | 24 |
| ENSG00000129270 | -1.35 | 1.69E-06 | *MMP28* | Protein coding | 24 |
| ENSG00000146592 | -1.35 | 5.42E-22 | *CREB5* | Protein coding | 24 |
| ENSG00000079482 | -1.35 | 1.65E-05 | *OPHN1* | Protein coding | 24 |
| ENSG00000100292 | 1.35 | 1.48E-12 | *HMOX1* | Protein coding | 24 |
| ENSG00000175003 | 1.35 | 9.81E-05 | *SLC22A1* | Protein coding | 24 |
| ENSG00000256574 | -1.35 | 0.00504 | *OR13A1* | Protein coding | 24 |
| ENSG00000003402 | 1.35 | 1.00E-62 | *CFLAR* | Protein coding | 24 |
| ENSG00000107551 | 1.35 | 6.98E-15 | *RASSF4* | Protein coding | 24 |
| ENSG00000044459 | 1.35 | 8.10E-10 | *CNTLN* | Protein coding | 24 |
| ENSG00000176014 | 1.35 | 1.78E-10 | *TUBB6* | Protein coding | 24 |
| ENSG00000114405 | 1.35 | 6.51E-05 | *C3orf14* | Protein coding | 24 |
| ENSG00000134107 | 1.35 | 6.35E-26 | *BHLHE40* | Protein coding | 24 |
| ENSG00000158966 | -1.35 | 0.00015 | *CACHD1* | Protein coding | 24 |
| ENSG00000100427 | -1.35 | 1.23E-08 | *MLC1* | Protein coding | 24 |
| ENSG00000038427 | 1.35 | 2.35E-05 | *VCAN* | Protein coding | 24 |
| ENSG00000204936 | 1.35 | 4.63E-06 | *CD177* | Protein coding | 24 |
| ENSG00000149489 | 1.35 | 2.33E-09 | *ROM1* | Protein coding | 24 |
| ENSG00000146094 | 1.34 | 6.96E-14 | *DOK3* | Protein coding | 24 |
| ENSG00000183307 | 1.34 | 8.88E-06 | *CECR6* | Protein coding | 24 |
| ENSG00000127920 | 1.34 | 9.84E-08 | *GNG11* | Protein coding | 24 |
| ENSG00000152760 | 1.34 | 0.00705 | *TCTEX1D1* | Protein coding | 24 |
| ENSG00000082397 | 1.34 | 4.16E-16 | *EPB41L3* | Protein coding | 24 |
| ENSG00000169554 | 1.34 | 4.44E-35 | *ZEB2* | Protein coding | 24 |
| ENSG00000118523 | 1.34 | 1.65E-06 | *CTGF* | Protein coding | 24 |
| ENSG00000169252 | -1.34 | 5.19E-22 | *ADRB2* | Protein coding | 24 |
| ENSG00000088298 | 1.34 | 5.29E-17 | *EDEM2* | Protein coding | 24 |
| ENSG00000102096 | 1.34 | 1.96E-34 | *PIM2* | Protein coding | 24 |
| ENSG00000090924 | 1.34 | 4.61E-23 | *PLEKHG2* | Protein coding | 24 |
| ENSG00000175643 | 1.34 | 3.25E-15 | *RMI2* | Protein coding | 24 |
| ENSG00000143217 | 1.34 | 0.00037 | *PVRL4* | Protein coding | 24 |
| ENSG00000066468 | -1.34 | 0.00050 | *FGFR2* | Protein coding | 24 |
| ENSG00000085449 | 1.33 | 5.03E-38 | *WDFY1* | Protein coding | 24 |
| ENSG00000155307 | 1.33 | 5.26E-24 | *SAMSN1* | Protein coding | 24 |
| ENSG00000204169 | -1.33 | 0.02536 | *AGAP7* | Protein coding | 24 |
| ENSG00000068366 | 1.33 | 1.93E-35 | *ACSL4* | Protein coding | 24 |
| ENSG00000241839 | 1.33 | 1.62E-19 | *PLEKHO2* | Protein coding | 24 |
| ENSG00000173559 | 1.33 | 2.30E-49 | *NABP1* | Protein coding | 24 |
| ENSG00000158517 | 1.33 | 2.37E-09 | *NCF1* | Protein coding | 24 |
| ENSG00000145687 | -1.33 | 2.30E-20 | *SSBP2* | Protein coding | 24 |
| ENSG00000198283 | 1.32 | 0.01903 | *OR5B21* | Protein coding | 24 |
| ENSG00000182557 | -1.32 | 0.00075 | *SPNS3* | Protein coding | 24 |
| ENSG00000198355 | 1.32 | 1.10E-11 | *PIM3* | Protein coding | 24 |
| ENSG00000007866 | 1.32 | 2.38E-06 | *TEAD3* | Protein coding | 24 |
| ENSG00000075399 | 1.32 | 7.39E-15 | *VPS9D1* | Protein coding | 24 |
| ENSG00000085733 | 1.32 | 0.00023 | *CTTN* | Protein coding | 24 |
| ENSG00000180071 | -1.32 | 2.65E-06 | *ANKRD18A* | Protein coding | 24 |
| ENSG00000184232 | 1.32 | 4.05E-10 | *OAF* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000161888 | -1.32 | 0.01022 | *SPC24* | Protein coding | 24 |
| ENSG00000100284 | 1.32 | 1.17E-14 | *TOM1* | Protein coding | 24 |
| ENSG00000147443 | -1.32 | 4.89E-25 | *DOK2* | Protein coding | 24 |
| ENSG00000173706 | 1.32 | 1.10E-27 | *HEG1* | Protein coding | 24 |
| ENSG00000182600 | -1.32 | 0.00109 | *C2orf82* | Protein coding | 24 |
| ENSG00000057294 | -1.32 | 0.02264 | *PKP2* | Protein coding | 24 |
| ENSG00000013619 | 1.31 | 2.37E-07 | *MAMLD1* | Protein coding | 24 |
| ENSG00000116701 | 1.31 | 4.01E-14 | *NCF2* | Protein coding | 24 |
| ENSG00000151117 | -1.31 | 8.65E-13 | *TMEM86A* | Protein coding | 24 |
| ENSG00000139193 | -1.31 | 1.34E-26 | *CD27* | Protein coding | 24 |
| ENSG00000173868 | -1.31 | 2.37E-09 | *PHOSPHO1* | Protein coding | 24 |
| ENSG00000146376 | 1.31 | 1.07E-16 | *ARHGAP18* | Protein coding | 24 |
| ENSG00000130830 | 1.31 | 2.65E-11 | *MPP1* | Protein coding | 24 |
| ENSG00000189292 | -1.31 | 0.02671 | *FAM150B* | Protein coding | 24 |
| ENSG00000085491 | 1.31 | 2.30E-29 | *SLC25A24* | Protein coding | 24 |
| ENSG00000132357 | 1.31 | 1.32E-20 | *CARD6* | Protein coding | 24 |
| ENSG00000025039 | 1.31 | 5.76E-21 | *RRAGD* | Protein coding | 24 |
| ENSG00000159128 | 1.31 | 2.18E-20 | *IFNGR2* | Protein coding | 24 |
| ENSG00000091436 | 1.31 | 3.21E-19 | *MLTK* | Protein coding | 24 |
| ENSG00000050030 | -1.31 | 1.91E-05 | *KIAA2022* | Protein coding | 24 |
| ENSG00000148154 | 1.30 | 1.24E-41 | *UGCG* | Protein coding | 24 |
| ENSG00000123609 | 1.30 | 4.62E-17 | *NMI* | Protein coding | 24 |
| ENSG00000176884 | 1.30 | 0.01711 | *GRIN1* | Protein coding | 24 |
| ENSG00000142192 | 1.30 | 1.29E-18 | *APP* | Protein coding | 24 |
| ENSG00000116991 | 1.30 | 6.36E-14 | *SIPA1L2* | Protein coding | 24 |
| ENSG00000143570 | 1.30 | 4.66E-20 | *SLC39A1* | Protein coding | 24 |
| ENSG00000196923 | 1.30 | 4.62E-16 | *PDLIM7* | Protein coding | 24 |
| ENSG00000135643 | -1.30 | 4.26E-05 | *KCNMB4* | Protein coding | 24 |
| ENSG00000140464 | 1.30 | 5.54E-10 | *PML* | Protein coding | 24 |
| ENSG00000206077 | -1.30 | 4.75E-14 | *ZDHHC11B* | Protein coding | 24 |
| ENSG00000172159 | 1.30 | 1.45E-09 | *FRMD3* | Protein coding | 24 |
| ENSG00000059378 | 1.30 | 2.45E-09 | *PARP12* | Protein coding | 24 |
| ENSG00000136161 | -1.29 | 2.98E-14 | *RCBTB2* | Protein coding | 24 |
| ENSG00000128040 | -1.29 | 0.03808 | *SPINK2* | Protein coding | 24 |
| ENSG00000104722 | -1.29 | 0.00021 | *NEFM* | Protein coding | 24 |
| ENSG00000091428 | 1.29 | 0.01203 | *RAPGEF4* | Protein coding | 24 |
| ENSG00000196782 | -1.29 | 2.60E-19 | *MAML3* | Protein coding | 24 |
| ENSG00000154640 | 1.29 | 1.37E-20 | *BTG3* | Protein coding | 24 |
| ENSG00000157985 | -1.29 | 3.92E-07 | *AGAP1* | Protein coding | 24 |
| ENSG00000163803 | -1.29 | 3.28E-10 | *PLB1* | Protein coding | 24 |
| ENSG00000121060 | 1.29 | 1.68E-15 | *TRIM25* | Protein coding | 24 |
| ENSG00000125733 | 1.29 | 9.98E-24 | *TRIP10* | Protein coding | 24 |
| ENSG00000174125 | 1.29 | 5.84E-39 | *TLR1* | Protein coding | 24 |
| ENSG00000137575 | 1.29 | 1.25E-31 | *SDCBP* | Protein coding | 24 |
| ENSG00000105479 | -1.29 | 0.00894 | *CCDC114* | Protein coding | 24 |
| ENSG00000144668 | 1.29 | 0.00316 | *ITGA9* | Protein coding | 24 |
| ENSG00000169994 | 1.28 | 9.41E-05 | *MYO7B* | Protein coding | 24 |
| ENSG00000180611 | 1.28 | 8.84E-15 | *MB21D2* | Protein coding | 24 |
| ENSG00000150681 | -1.28 | 1.82E-10 | *RGS18* | Protein coding | 24 |
| ENSG00000105967 | 1.28 | 9.67E-24 | *TFEC* | Protein coding | 24 |
| ENSG00000102837 | 1.28 | 8.60E-11 | *OLFM4* | Protein coding | 24 |
| ENSG00000197614 | 1.28 | 0.03890 | *MFAP5* | Protein coding | 24 |
| ENSG00000127951 | -1.28 | 0.00034 | *FGL2* | Protein coding | 24 |
| ENSG00000173578 | 1.28 | 9.43E-05 | *XCR1* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000119535 | -1.28 | 6.24E-13 | *CSF3R* | Protein coding | 24 |
| ENSG00000085831 | 1.28 | 0.00980 | *TTC39A* | Protein coding | 24 |
| ENSG00000176092 | -1.28 | 3.71E-05 | *AIM1L* | Protein coding | 24 |
| ENSG00000167291 | 1.28 | 5.04E-16 | *TBC1D16* | Protein coding | 24 |
| ENSG00000138678 | -1.28 | 1.96E-11 | *AGPAT9* | Protein coding | 24 |
| ENSG00000162894 | -1.28 | 1.52E-36 | *FAIM3* | Protein coding | 24 |
| ENSG00000076067 | -1.28 | 1.58E-07 | *RBMS2* | Protein coding | 24 |
| ENSG00000137731 | -1.28 | 0.03045 | *FXYD2* | Protein coding | 24 |
| ENSG00000102678 | -1.28 | 4.03E-06 | *FGF9* | Protein coding | 24 |
| ENSG00000039319 | 1.28 | 1.10E-21 | *ZFYVE16* | Protein coding | 24 |
| ENSG00000155096 | 1.28 | 9.24E-35 | *AZIN1* | Protein coding | 24 |
| ENSG00000186815 | -1.28 | 5.39E-37 | *TPCN1* | Protein coding | 24 |
| ENSG00000075426 | 1.28 | 8.13E-16 | *FOSL2* | Protein coding | 24 |
| ENSG00000141968 | 1.27 | 8.47E-31 | *VAV1* | Protein coding | 24 |
| ENSG00000004838 | -1.27 | 5.85E-08 | *ZMYND10* | Protein coding | 24 |
| ENSG00000100647 | 1.27 | 2.19E-27 | *KIAA0247* | Protein coding | 24 |
| ENSG00000107242 | -1.27 | 4.59E-08 | *PIP5K1B* | Protein coding | 24 |
| ENSG00000176845 | 1.27 | 3.76E-13 | *METRNL* | Protein coding | 24 |
| ENSG00000073792 | 1.27 | 1.16E-11 | *IGF2BP2* | Protein coding | 24 |
| ENSG00000167549 | -1.27 | 2.70E-06 | *CORO6* | Protein coding | 24 |
| ENSG00000169744 | -1.27 | 0.00146 | *LDB2* | Protein coding | 24 |
| ENSG00000154240 | 1.27 | 2.29E-06 | *CEP112* | Protein coding | 24 |
| ENSG00000130203 | -1.27 | 0.0115 | *APOE* | Protein coding | 24 |
| ENSG00000178038 | -1.27 | 5.78E-12 | *ALS2CL* | Protein coding | 24 |
| ENSG00000105808 | -1.27 | 0.00053 | *RASA4* | Protein coding | 24 |
| ENSG00000049449 | 1.27 | 3.16E-11 | *RCN1* | Protein coding | 24 |
| ENSG00000106571 | 1.27 | 0.02524 | *GLI3* | Protein coding | 24 |
| ENSG00000043462 | 1.27 | 7.03E-30 | *LCP2* | Protein coding | 24 |
| ENSG00000136274 | 1.26 | 0.01974 | *NACAD* | Protein coding | 24 |
| ENSG00000184014 | 1.26 | 7.83E-42 | *DENND5A* | Protein coding | 24 |
| ENSG00000143409 | -1.26 | 1.58E-21 | *FAM63A* | Protein coding | 24 |
| ENSG00000159403 | 1.26 | 0.00016 | *C1R* | Protein coding | 24 |
| ENSG00000014257 | -1.26 | 9.52E-08 | *ACPP* | Protein coding | 24 |
| ENSG00000064763 | 1.26 | 4.48E-14 | *FAR2* | Protein coding | 24 |
| ENSG00000145103 | 1.26 | 0.00525 | *ILDR1* | Protein coding | 24 |
| ENSG00000167768 | -1.26 | 2.16E-09 | *KRT1* | Protein coding | 24 |
| ENSG00000136929 | -1.26 | 7.16E-06 | *HEMGN* | Protein coding | 24 |
| ENSG00000108984 | -1.26 | 6.51E-10 | *MAP2K6* | Protein coding | 24 |
| ENSG00000113657 | 1.26 | 0.00777 | *DPYSL3* | Protein coding | 24 |
| ENSG00000183735 | 1.26 | 4.28E-48 | *TBK1* | Protein coding | 24 |
| ENSG00000168404 | 1.26 | 3.39E-14 | *MLKL* | Protein coding | 24 |
| ENSG00000101294 | 1.25 | 8.32E-16 | *HM13* | Protein coding | 24 |
| ENSG00000113742 | 1.25 | 2.09E-41 | *CPEB4* | Protein coding | 24 |
| ENSG00000165259 | 1.25 | 1.38E-05 | *HDX* | Protein coding | 24 |
| ENSG00000137496 | 1.25 | 2.84E-08 | *IL18BP* | Protein coding | 24 |
| ENSG00000165806 | 1.25 | 9.85E-15 | *CASP7* | Protein coding | 24 |
| ENSG00000135378 | 1.25 | 3.49E-07 | *PRRG4* | Protein coding | 24 |
| ENSG00000132141 | -1.25 | 0.00434 | *CCT6B* | Protein coding | 24 |
| ENSG00000108861 | 1.25 | 1.97E-15 | *DUSP3* | Protein coding | 24 |
| ENSG00000132334 | 1.25 | 3.23E-20 | *PTPRE* | Protein coding | 24 |
| ENSG00000145708 | -1.25 | 0.03546 | *CRHBP* | Protein coding | 24 |
| ENSG00000008513 | 1.25 | 5.07E-18 | *ST3GAL1* | Protein coding | 24 |
| ENSG00000168994 | 1.25 | 4.48E-08 | *PXDC1* | Protein coding | 24 |
| ENSG00000103043 | 1.25 | 7.33E-25 | *VAC14* | Protein coding | 24 |

| Ensembl ID | Log2Fold Change | P adjusted | Gene | Biotype | Stimulation time (hours) |
|---|---|---|---|---|---|
| ENSG00000204267 | 1.25 | 6.53E-18 | *TAP2* | Protein coding | 24 |
| ENSG00000134668 | -1.25 | 0.00100 | *SPOCD1* | Protein coding | 24 |
| ENSG00000072274 | 1.25 | 1.03E-38 | *TFRC* | Protein coding | 24 |
| ENSG00000164430 | 1.25 | 6.56E-16 | *MB21D1* | Protein coding | 24 |
| ENSG00000146859 | 1.24 | 4.33E-15 | *TMEM140* | Protein coding | 24 |
| ENSG00000141540 | -1.24 | 4.92E-19 | *TTYH2* | Protein coding | 24 |
| ENSG00000116260 | 1.24 | 3.27E-11 | *QSOX1* | Protein coding | 24 |
| ENSG00000090674 | 1.24 | 7.38E-12 | *MCOLN1* | Protein coding | 24 |
| ENSG00000188004 | 1.24 | 5.47E-08 | *C1orf204* | Protein coding | 24 |
| ENSG00000168229 | -1.24 | 4.86E-13 | *PTGDR* | Protein coding | 24 |
| ENSG00000152672 | -1.24 | 0.04704 | *CLEC4F* | Protein coding | 24 |
| ENSG00000111321 | 1.24 | 3.81E-08 | *LTBR* | Protein coding | 24 |
| ENSG00000198363 | 1.24 | 3.53E-23 | *ASPH* | Protein coding | 24 |
| ENSG00000170293 | 1.24 | 4.06E-14 | *CMTM8* | Protein coding | 24 |
| ENSG00000111224 | 1.24 | 2.82E-22 | *PARP11* | Protein coding | 24 |
| ENSG00000137841 | -1.24 | 3.33E-64 | *PLCB2* | Protein coding | 24 |
| ENSG00000090975 | -1.24 | 4.19E-24 | *PITPNM2* | Protein coding | 24 |
| ENSG00000212743 | 1.24 | 6.58E-08 | *DKFZP667F0711* | Protein coding | 24 |
| ENSG00000137507 | 1.24 | 1.44E-12 | *LRRC32* | Protein coding | 24 |
| ENSG00000007062 | 1.24 | 0.01018 | *PROM1* | Protein coding | 24 |
| ENSG00000185567 | -1.24 | 0.00016 | *AHNAK2* | Protein coding | 24 |
| ENSG00000147883 | 1.23 | 1.48E-07 | *CDKN2B* | Protein coding | 24 |
| ENSG00000136560 | 1.23 | 4.60E-45 | *TANK* | Protein coding | 24 |
| ENSG00000198964 | 1.23 | 7.81E-26 | *SGMS1* | Protein coding | 24 |
| ENSG00000013364 | 1.23 | 6.26E-13 | *MVP* | Protein coding | 24 |
| ENSG00000164342 | 1.23 | 2.96E-09 | *TLR3* | Protein coding | 24 |
| ENSG00000086065 | 1.23 | 1.12E-17 | *CHMP5* | Protein coding | 24 |
| ENSG00000119139 | 1.23 | 6.70E-15 | *TJP2* | Protein coding | 24 |
| ENSG00000106537 | 1.23 | 2.13E-17 | *TSPAN13* | Protein coding | 24 |
| ENSG00000143507 | 1.23 | 3.38E-27 | *DUSP10* | Protein coding | 24 |
| ENSG00000109861 | 1.23 | 0.00143 | *CTSC* | Protein coding | 24 |
| ENSG00000269089 | 1.23 | 0.01105 | *AP003733.1* | Protein coding | 24 |
| ENSG00000080573 | -1.23 | 3.89E-10 | *COL5A3* | Protein coding | 24 |
| ENSG00000115112 | -1.23 | 6.76E-06 | *TFCP2L1* | Protein coding | 24 |
| ENSG00000121858 | 1.23 | 0.00133 | *TNFSF10* | Protein coding | 24 |
| ENSG00000204131 | -1.23 | 8.46E-15 | *NHSL2* | Protein coding | 24 |
| ENSG00000113296 | -1.23 | 3.17E-07 | *THBS4* | Protein coding | 24 |

**Table S2.** A total of 246 protein-coding genes were differentially expressed at both 4 and 24 hour-*Candida* stimulation, showing >1.5-fold higher expression compared to RPMI medium used as control.

| Ensembl ID | Gene | Ensembl ID | Gene |
|---|---|---|---|
| ENSG00000154262 | *ABCA6* | ENSG00000135678 | *CPM* |
| ENSG00000166016 | *ABTB2* | ENSG00000164400 | *CSF2* |
| ENSG00000217825 | *AC099552.4* | ENSG00000108342 | *CSF3* |
| ENSG00000151726 | *ACSL1* | ENSG00000163739 | *CXCL1* |
| ENSG00000148926 | *ADM* | ENSG00000169245 | *CXCL10* |
| ENSG00000162433 | *AK4* | ENSG00000081041 | *CXCL2* |
| ENSG00000215306 | *AL135998.1* | ENSG00000163734 | *CXCL3* |
| ENSG00000114019 | *AMOTL2* | ENSG00000163735 | *CXCL5* |
| ENSG00000128383 | *APOBEC3A* | ENSG00000124875 | *CXCL6* |
| ENSG00000103569 | *AQP9* | ENSG00000137869 | *CYP19A1* |
| ENSG00000162772 | *ATF3* | ENSG00000134716 | *CYP2J2* |
| ENSG00000158470 | *B4GALT5* | ENSG00000107201 | *DDX58* |
| ENSG00000156127 | *BATF* | ENSG00000137628 | *DDX60* |
| ENSG00000168062 | *BATF2* | ENSG00000164825 | *DEFB1* |
| ENSG00000123685 | *BATF3* | ENSG00000105928 | *DFNA5* |
| ENSG00000064787 | *BCAS1* | ENSG00000108771 | *DHX58* |
| ENSG00000140379 | *BCL2A1* | ENSG00000136048 | *DRAM1* |
| ENSG00000121380 | *BCL2L14* | ENSG00000163840 | *DTX3L* |
| ENSG00000187479 | *C11orf96* | ENSG00000138166 | *DUSP5* |
| ENSG00000166920 | *C15orf48* | ENSG00000105246 | *EBI3* |
| ENSG00000197982 | *C1orf122* | ENSG00000078401 | *EDN1* |
| ENSG00000125462 | *C1orf61* | ENSG00000135373 | *EHF* |
| ENSG00000173918 | *C1QTNF1* | ENSG00000055332 | *EIF2AK2* |
| ENSG00000125730 | *C3* | ENSG00000164181 | *ELOVL7* |
| ENSG00000141837 | *CACNA1A* | ENSG00000133106 | *EPSTI1* |
| ENSG00000008118 | *CAMK1G* | ENSG00000157557 | *ETS2* |
| ENSG00000137757 | *CASP5* | ENSG00000253831 | *ETV3L* |
| ENSG00000108691 | *CCL2* | ENSG00000010030 | *ETV7* |
| ENSG00000115009 | *CCL20* | ENSG00000117525 | *F3* |
| ENSG00000167236 | *CCL23* | ENSG00000198673 | *FAM19A2* |
| ENSG00000006075 | *CCL3* | ENSG00000026103 | *FAS* |
| ENSG00000205021 | *CCL3L1* | ENSG00000116663 | *FBXO6* |
| ENSG00000129277 | *CCL4* | ENSG00000149557 | *FEZ1* |
| ENSG00000205020 | *CCL4L1* | ENSG00000126262 | *FFAR2* |
| ENSG00000197262 | *CCL4L2* | ENSG00000179431 | *FJX1* |
| ENSG00000108688 | *CCL7* | ENSG00000102755 | *FLT1* |
| ENSG00000108700 | *CCL8* | ENSG00000075618 | *FSCN1* |
| ENSG00000101331 | *CCM2L* | ENSG00000106701 | *FSD1L* |
| ENSG00000121797 | *CCRL2* | ENSG00000123689 | *G0S2* |
| ENSG00000120217 | *CD274* | ENSG00000130222 | *GADD45G* |
| ENSG00000121594 | *CD80* | ENSG00000131979 | *GCH1* |
| ENSG00000149798 | *CDC42EP2* | ENSG00000121743 | *GJA3* |
| ENSG00000198848 | *CES1* | ENSG00000165474 | *GJB2* |
| ENSG00000243649 | *CFB* | ENSG00000139572 | *GPR84* |
| ENSG00000138135 | *CH25H* | ENSG00000013588 | *GPRC5A* |
| ENSG00000114737 | *CISH* | ENSG00000180875 | *GREM2* |
| ENSG00000166523 | *CLEC4E* | ENSG00000182782 | *HCAR2* |
| ENSG00000153132 | *CLGN* | ENSG00000255398 | *HCAR3* |
| ENSG00000169504 | *CLIC4* | ENSG00000130589 | *HELZ2* |
| ENSG00000134326 | *CMPK2* | ENSG00000138646 | *HERC5* |

| Ensembl ID | Gene | Ensembl ID | Gene | Ensembl ID | Gene |
|---|---|---|---|---|---|
| ENSG00000138642 | *HERC6* | ENSG00000168389 | *MFSD2A* | ENSG00000124107 | *SLPI* |
| ENSG00000164683 | *HEY1* | ENSG00000074416 | *MGLL* | ENSG00000214872 | *SMTNL1* |
| ENSG00000105707 | *HPN* | ENSG00000166670 | *MMP10* | ENSG00000185338 | *SOCS1* |
| ENSG00000196639 | *HRH1* | ENSG00000157227 | *MMP14* | ENSG00000184557 | *SOCS3* |
| ENSG00000117594 | *HSD11B1* | ENSG00000178860 | *MSC* | ENSG00000112096 | *SOD2* |
| ENSG00000090339 | *ICAM1* | ENSG00000125148 | *MT2A* | ENSG00000186583 | *SPATC1* |
| ENSG00000137331 | *IER3* | ENSG00000157601 | *MX1* | ENSG00000196141 | *SPATS2L* |
| ENSG00000165949 | *IFI27* | ENSG00000183486 | *MX2* | ENSG00000164266 | *SPINK1* |
| ENSG00000068079 | *IFI35* | ENSG00000104320 | *NBN* | ENSG00000180616 | *SSTR2* |
| ENSG00000137965 | *IFI44* | ENSG00000204099 | *NEU4* | ENSG00000183473 | *SSTR3* |
| ENSG00000137959 | *IFI44L* | ENSG00000162614 | *NEXN* | ENSG00000172403 | *SYNPO2* |
| ENSG00000126709 | *IFI6* | ENSG00000100906 | *NFKBIA* | ENSG00000148737 | *TCF7L2* |
| ENSG00000115267 | *IFIH1* | ENSG00000144802 | *NFKBIZ* | ENSG00000196116 | *TDRD7* |
| ENSG00000185745 | *IFIT1* | ENSG00000131669 | *NINJ1* | ENSG00000105825 | *TFPI2* |
| ENSG00000119922 | *IFIT2* | ENSG00000089127 | *OAS1* | ENSG00000137462 | *TLR2* |
| ENSG00000119917 | *IFIT3* | ENSG00000111335 | *OAS2* | ENSG00000121900 | *TMEM54* |
| ENSG00000152778 | *IFIT5* | ENSG00000111331 | *OAS3* | ENSG00000232810 | *TNF* |
| ENSG00000185885 | *IFITM1* | ENSG00000135114 | *OASL* | ENSG00000185215 | *TNFAIP2* |
| ENSG00000142089 | *IFITM3* | ENSG00000184221 | *OLIG1* | ENSG00000123610 | *TNFAIP6* |
| ENSG00000111537 | *IFNG* | ENSG00000205927 | *OLIG2* | ENSG00000121858 | *TNFSF10* |
| ENSG00000136634 | *IL10* | ENSG00000171631 | *P2RY6* | ENSG00000181634 | *TNFSF15* |
| ENSG00000113302 | *IL12B* | ENSG00000059378 | *PARP12* | ENSG00000050730 | *TNIP3* |
| ENSG00000169194 | *IL13* | ENSG00000138496 | *PARP9* | ENSG00000132109 | *TRIM21* |
| ENSG00000134470 | *IL15RA* | ENSG00000162493 | *PDPN* | ENSG00000132274 | *TRIM22* |
| ENSG00000142224 | *IL19* | ENSG00000170525 | *PFKFB3* | ENSG00000152503 | *TRIM36* |
| ENSG00000115008 | *IL1A* | ENSG00000181649 | *PHLDA2* | ENSG00000125733 | *TRIP10* |
| ENSG00000125538 | *IL1B* | ENSG00000124102 | *PI3* | ENSG00000136810 | *TXN* |
| ENSG00000136689 | *IL1RN* | ENSG00000102096 | *PIM2* | ENSG00000156587 | *UBE2L6* |
| ENSG00000197272 | *IL27* | ENSG00000122861 | *PLAU* | ENSG00000100024 | *UPB1* |
| ENSG00000134460 | *IL2RA* | ENSG00000011422 | *PLAUR* | ENSG00000184979 | *USP18* |
| ENSG00000136688 | *IL36G* | ENSG00000075651 | *PLD1* | ENSG00000114251 | *WNT5A* |
| ENSG00000136695 | *IL36RN* | ENSG00000115956 | *PLEK* | ENSG00000132530 | *XAF1* |
| ENSG00000104951 | *IL4I1* | ENSG00000188313 | *PLSCR1* | ENSG00000168334 | *XIRP1* |
| ENSG00000136244 | *IL6* | ENSG00000140464 | *PML* | ENSG00000149289 | *ZC3H12C* |
| ENSG00000104432 | *IL7* | ENSG00000180316 | *PNPLA1* | | |
| ENSG00000169429 | *IL8* | ENSG00000183657 | *PP13439* | | |
| ENSG00000122641 | *INHBA* | ENSG00000087074 | *PPP1R15A* | | |
| ENSG00000134070 | *IRAK2* | ENSG00000152229 | *PSTPIP2* | | |
| ENSG00000185507 | *IRF7* | ENSG00000073756 | *PTGS2* | | |
| ENSG00000102794 | *IRG1* | ENSG00000163661 | *PTX3* | | |
| ENSG00000187608 | *ISG15* | ENSG00000143344 | *RGL1* | | |
| ENSG00000172183 | *ISG20* | ENSG00000132669 | *RIN2* | | |
| ENSG00000105855 | *ITGB8* | ENSG00000104312 | *RIPK2* | | |
| ENSG00000123700 | *KCNJ2* | ENSG00000137393 | *RNF144B* | | |
| ENSG00000102554 | *KLF5* | ENSG00000235531 | *RP11-383H13.1* | | |
| ENSG00000117009 | *KMO* | ENSG00000134321 | *RSAD2* | | |
| ENSG00000089692 | *LAG3* | ENSG00000136514 | *RTP4* | | |
| ENSG00000196878 | *LAMB3* | ENSG00000177409 | *SAMD9L* | | |
| ENSG00000078081 | *LAMP3* | ENSG00000105711 | *SCN1B* | | |
| ENSG00000198121 | *LPAR1* | ENSG00000197632 | *SERPINB2* | | |
| ENSG00000160932 | *LY6E* | ENSG00000163082 | *SGPP2* | | |
| ENSG00000183742 | *MACC1* | ENSG00000088827 | *SIGLEC1* | | |
| ENSG00000185022 | *MAFF* | ENSG00000026751 | *SLAMF7* | | |
| ENSG00000107968 | *MAP3K8* | ENSG00000079215 | *SLC1A3* | | |
| ENSG00000155130 | *MARCKS* | ENSG00000160326 | *SLC2A6* | | |
| ENSG00000175471 | *MCTP1* | ENSG00000138821 | *SLC39A8* | | |

**Table S3.** Pathway enrichment analysis on differentially expressed protein-coding genes that showed >1.5 higher expression compared to RPMI medium in response to 4 hour-*Candida* stimulation.

| Pathway name | Set size | Candidates | p-value | q-value | Pathway source | -log10P |
|---|---|---|---|---|---|---|
| Immune System | 1174 | 60 (5.2%) | 4.77E-13 | 5.33E-11 | Reactome | 12.32 |
| Cytokine Signaling in Immune system | 376 | 32 (8.5%) | 1.36E-12 | 1.06E-10 | Reactome | 11.87 |
| Interferon Signaling | 68 | 13 (19.1%) | 4.90E-10 | 2.74E-08 | Reactome | 9.31 |
| Chemokine receptors bind chemokines | 60 | 11 (19.0%) | 1.20E-08 | 4.09E-07 | Reactome | 7.92 |
| ISG15 antiviral mechanism | 31 | 8 (25.8%) | 9.71E-08 | 2.00E-06 | Reactome | 7.01 |
| Antiviral mechanism by IFN-stimulated genes | 31 | 8 (25.8%) | 9.71E-08 | 2.00E-06 | Reactome | 7.01 |
| Class A/1 (Rhodopsin-like receptors) | 326 | 22 (6.8%) | 3.00E-07 | 5.23E-06 | Reactome | 6.52 |
| Innate Immune System | 709 | 32 (4.6%) | 5.44E-06 | 8.20E-05 | Reactome | 5.26 |
| GPCR ligand binding | 454 | 24 (5.4%) | 6.55E-06 | 9.49E-05 | Reactome | 5.18 |
| RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways | 53 | 8 (15.1%) | 7.55E-06 | 0.000107 | Reactome | 5.12 |
| Peptide ligand-binding receptors | 199 | 14 (7.2%) | 2.69E-05 | 0.000352 | Reactome | 4.57 |
| Negative regulators of RIG-I/MDA5 signaling | 21 | 5 (23.8%) | 4.21E-05 | 0.000515 | Reactome | 4.38 |
| Signaling by Interleukins | 270 | 16 (5.9%) | 7.20E-05 | 0.000829 | Reactome | 4.14 |
| Interleukin-1 signaling | 43 | 6 (14.0%) | 0.000169 | 0.00177 | Reactome | 3.77 |
| Interleukin-1 processing | 7 | 3 (42.9%) | 0.00024 | 0.00238 | Reactome | 3.62 |
| STING mediated induction of host immune responses | 17 | 4 (23.5%) | 0.000272 | 0.00263 | Reactome | 3.57 |
| Cytosolic sensors of pathogen-associated DNA | 49 | 6 (12.2%) | 0.000353 | 0.00325 | Reactome | 3.45 |
| G alpha (i) signalling events | 243 | 13 (5.4%) | 0.000834 | 0.00718 | Reactome | 3.08 |
| Hydroxycarboxylic acid-binding receptors | 3 | 2 (66.7%) | 0.00112 | 0.0093 | Reactome | 2.95 |
| Regulation of IFNA signaling | 26 | 4 (15.4%) | 0.00149 | 0.012 | Reactome | 2.83 |
| Interferon alpha/beta signaling | 27 | 4 (14.8%) | 0.00172 | 0.0138 | Reactome | 2.76 |
| TRAF3-dependent IRF activation pathway | 13 | 3 (23.1%) | 0.0018 | 0.0141 | Reactome | 2.74 |
| Dissolution of Fibrin Clot | 13 | 3 (23.1%) | 0.0018 | 0.0141 | Reactome | 2.74 |
| Alternative complement activation | 4 | 2 (50.0%) | 0.0022 | 0.0165 | Reactome | 2.66 |
| Regulation of IFNG signaling | 14 | 3 (21.4%) | 0.00226 | 0.0165 | Reactome | 2.65 |
| IRF3-mediated induction of type I IFN | 14 | 3 (21.4%) | 0.00226 | 0.0165 | Reactome | 2.65 |
| Regulation of innate immune responses to cytosolic DNA | 16 | 3 (18.8%) | 0.00338 | 0.0232 | Reactome | 2.47 |
| TRAF6 mediated IRF7 activation | 17 | 3 (17.6%) | 0.00404 | 0.0266 | Reactome | 2.39 |
| MyD88 dependent cascade initiated on endosome | 79 | 6 (7.6%) | 0.00433 | 0.0275 | Reactome | 2.36 |
| Toll Like Receptor 7/8 (TLR7/8) Cascade | 79 | 6 (7.6%) | 0.00433 | 0.0275 | Reactome | 2.36 |
| Toll Like Receptor 9 (TLR9) Cascade | 83 | 6 (7.2%) | 0.00551 | 0.0334 | Reactome | 2.26 |
| Toll-Like Receptors Cascades | 138 | 8 (5.8%) | 0.00553 | 0.0334 | Reactome | 2.26 |
| Growth hormone receptor signaling | 19 | 3 (15.8%) | 0.0056 | 0.0334 | Reactome | 2.25 |
| Interferon gamma signaling | 19 | 3 (15.8%) | 0.0056 | 0.0334 | Reactome | 2.25 |
| Activated TLR4 signalling | 110 | 7 (6.4%) | 0.00562 | 0.0334 | Reactome | 2.25 |
| Adaptive Immune System | 612 | 21 (3.5%) | 0.00635 | 0.0368 | Reactome | 2.2 |
| MyD88:Mal cascade initiated on plasma membrane | 86 | 6 (7.0%) | 0.00655 | 0.0369 | Reactome | 2.18 |
| Toll Like Receptor TLR1:TLR2 Cascade | 86 | 6 (7.0%) | 0.00655 | 0.0369 | Reactome | 2.18 |
| Toll Like Receptor TLR6:TLR2 Cascade | 86 | 6 (7.0%) | 0.00655 | 0.0369 | Reactome | 2.18 |
| Toll Like Receptor 2 (TLR2) Cascade | 86 | 6 (7.0%) | 0.00655 | 0.0369 | Reactome | 2.18 |
| Activation of C3 and C5 | 7 | 2 (28.6%) | 0.00742 | 0.0412 | Reactome | 2.13 |
| Death Receptor Signalling | 22 | 3 (13.6%) | 0.00852 | 0.0462 | Reactome | 2.07 |
| Toll Like Receptor 4 (TLR4) Cascade | 119 | 7 (5.9%) | 0.00855 | 0.0462 | Reactome | 2.07 |
| RAF-independent MAPK1/3 activation | 23 | 3 (13.0%) | 0.00966 | 0.0504 | Reactome | 2.02 |

**Table S4.** Pathway enrichment analysis on differentially expressed protein-coding genes that showed >1.5 higher expression compared to RPMI medium in response to 24 hour-*Candida* stimulation.

| Pathway name | Set size | Candidates | p-value | q-value | Pathway source | -log10P |
|---|---|---|---|---|---|---|
| Chemokine receptors bind chemokines | 60 | 19 (32.8%) | 9.62E-11 | 3.04E-08 | Reactome | 10.02 |
| Immune System | 1174 | 113 (9.8%) | 2.42E-10 | 6.38E-08 | Reactome | 9.62 |
| Cytokine Signaling in Immune system | 376 | 50 (13.3%) | 3.19E-09 | 4.99E-07 | Reactome | 8.5 |
| Class A/1 (Rhodopsin-like receptors) | 326 | 44 (13.7%) | 1.39E-08 | 1.29E-06 | Reactome | 7.86 |
| Extracellular matrix organization | 264 | 38 (14.4%) | 3.09E-08 | 2.47E-06 | Reactome | 7.51 |
| Peptide ligand-binding receptors | 199 | 31 (15.9%) | 6.55E-08 | 4.93E-06 | Reactome | 7.18 |
| GPCR ligand binding | 454 | 51 (11.4%) | 4.05E-07 | 2.13E-05 | Reactome | 6.39 |
| Degradation of the extracellular matrix | 84 | 18 (21.4%) | 4.58E-07 | 2.33E-05 | Reactome | 6.34 |
| G alpha (i) signalling events | 243 | 33 (13.8%) | 8.08E-07 | 3.99E-05 | Reactome | 6.09 |
| HS-GAG biosynthesis | 31 | 10 (32.3%) | 3.37E-06 | 0.000137 | Reactome | 5.47 |
| Collagen degradation | 33 | 10 (30.3%) | 6.38E-06 | 0.000229 | Reactome | 5.2 |
| Hemostasis | 493 | 51 (10.4%) | 6.51E-06 | 0.000229 | Reactome | 5.19 |
| Non-integrin membrane-ECM interactions | 42 | 11 (26.2%) | 1.05E-05 | 0.000326 | Reactome | 4.98 |
| Interferon Signaling | 68 | 14 (20.6%) | 1.41E-05 | 0.000424 | Reactome | 4.85 |
| Signaling by Interleukins | 270 | 32 (11.9%) | 2.57E-05 | 0.000699 | Reactome | 4.59 |
| Innate Immune System | 709 | 64 (9.1%) | 2.90E-05 | 0.000765 | Reactome | 4.54 |
| Growth hormone receptor signaling | 19 | 7 (36.8%) | 3.92E-05 | 0.000954 | Reactome | 4.41 |
| HS-GAG degradation | 21 | 7 (33.3%) | 8.23E-05 | 0.00173 | Reactome | 4.08 |
| Cell surface interactions at the vascular wall | 101 | 16 (16.0%) | 9.35E-05 | 0.00194 | Reactome | 4.03 |
| Assembly of collagen fibrils and other multimeric structures | 44 | 10 (22.7%) | 9.92E-05 | 0.00204 | Reactome | 4 |
| Heparan sulfate/heparin (HS-GAG) metabolism | 54 | 11 (20.4%) | 0.000129 | 0.00256 | Reactome | 3.89 |
| Activation of Matrix Metalloproteinases | 30 | 8 (26.7%) | 0.00015 | 0.00286 | Reactome | 3.82 |
| Tryptophan catabolism | 11 | 5 (45.5%) | 0.000166 | 0.00306 | Reactome | 3.78 |
| ISG15 antiviral mechanism | 31 | 8 (25.8%) | 0.000192 | 0.00338 | Reactome | 3.72 |
| Antiviral mechanism by IFN-stimulated genes | 31 | 8 (25.8%) | 0.000192 | 0.00338 | Reactome | 3.72 |
| Gastrin-CREB signalling pathway via PKC and MAPK | 382 | 38 (10.0%) | 0.000222 | 0.00362 | Reactome | 3.65 |
| Collagen formation | 87 | 14 (16.1%) | 0.000238 | 0.00376 | Reactome | 3.62 |
| Syndecan interactions | 20 | 6 (30.0%) | 0.000515 | 0.00726 | Reactome | 3.29 |
| Astrocytic Glutamate-Glutamine Uptake And Metabolism | 4 | 3 (75.0%) | 0.000619 | 0.00843 | Reactome | 3.21 |
| Neurotransmitter uptake and Metabolism In Glial Cells | 4 | 3 (75.0%) | 0.000619 | 0.00843 | Reactome | 3.21 |
| Regulation of IFNG signaling | 14 | 5 (35.7%) | 0.000629 | 0.00849 | Reactome | 3.2 |
| Glycosaminoglycan metabolism | 119 | 16 (13.6%) | 0.000651 | 0.00872 | Reactome | 3.19 |
| SLC-mediated transmembrane transport | 263 | 27 (10.3%) | 0.00108 | 0.0133 | Reactome | 2.97 |
| Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell | 146 | 17 (12.5%) | 0.00114 | 0.0139 | Reactome | 2.94 |
| Laminin interactions | 23 | 6 (26.1%) | 0.00117 | 0.0141 | Reactome | 2.93 |
| Binding and Uptake of Ligands by Scavenger Receptors | 40 | 8 (20.0%) | 0.00122 | 0.0146 | Reactome | 2.91 |
| Dectin-2 family | 10 | 4 (40.0%) | 0.00141 | 0.0164 | Reactome | 2.85 |
| Platelet activation, signaling and aggregation | 229 | 24 (10.5%) | 0.0016 | 0.0181 | Reactome | 2.8 |
| Transport of inorganic cations/anions and amino acids/oligopeptides | 95 | 13 (13.7%) | 0.00187 | 0.021 | Reactome | 2.73 |
| Axon guidance | 459 | 40 (8.8%) | 0.00202 | 0.0219 | Reactome | 2.69 |
| Signaling by NOTCH3 | 11 | 4 (36.4%) | 0.00212 | 0.0228 | Reactome | 2.67 |
| Regulation of IFNA signaling | 26 | 6 (23.1%) | 0.00231 | 0.0244 | Reactome | 2.64 |
| A tetrasaccharide linker sequence is required for GAG synthesis | 26 | 6 (23.1%) | 0.00231 | 0.0244 | Reactome | 2.64 |
| VEGFR2 mediated cell proliferation | 198 | 21 (10.7%) | 0.00248 | 0.026 | Reactome | 2.61 |
| MAPK family signaling cascades | 225 | 23 (10.3%) | 0.00258 | 0.0266 | Reactome | 2.59 |
| Interferon alpha/beta signaling | 27 | 6 (22.2%) | 0.00284 | 0.0291 | Reactome | 2.55 |
| Scavenging by Class A Receptors | 19 | 5 (26.3%) | 0.00291 | 0.0295 | Reactome | 2.54 |
| Interferon gamma signaling | 19 | 5 (26.3%) | 0.00291 | 0.0295 | Reactome | 2.54 |

| Pathway name | Set size | Candidates | p-value | q-value | Pathway source | -log10P |
|---|---|---|---|---|---|---|
| MAPK1/MAPK3 signaling | 191 | 20 (10.5%) | 0.0036 | 0.0353 | Reactome | 2.44 |
| Signaling by Leptin | 193 | 20 (10.4%) | 0.00405 | 0.0389 | Reactome | 2.39 |
| Dissolution of Fibrin Clot | 13 | 4 (30.8%) | 0.00422 | 0.039 | Reactome | 2.37 |
| Negative regulators of RIG-I/MDA5 signaling | 21 | 5 (23.8%) | 0.00465 | 0.0422 | Reactome | 2.33 |
| Activation of C3 and C5 | 7 | 3 (42.9%) | 0.00479 | 0.0426 | Reactome | 2.32 |
| Anchoring fibril formation | 7 | 3 (42.9%) | 0.00479 | 0.0426 | Reactome | 2.32 |
| Signaling by SCF-KIT | 264 | 25 (9.5%) | 0.0048 | 0.0426 | Reactome | 2.32 |
| NCAM signaling for neurite out-growth | 223 | 22 (9.9%) | 0.00485 | 0.0428 | Reactome | 2.31 |
| Chondroitin sulfate/dermatan sulfate metabolism | 51 | 8 (16.0%) | 0.00526 | 0.0455 | Reactome | 2.28 |
| Interleukin-3, 5 and GM-CSF signaling | 211 | 21 (10.0%) | 0.00527 | 0.0455 | Reactome | 2.28 |
| RAF/MAP kinase cascade | 185 | 19 (10.3%) | 0.00551 | 0.0456 | Reactome | 2.26 |
| SHC1 events in EGFR signaling | 185 | 19 (10.3%) | 0.00551 | 0.0456 | Reactome | 2.26 |
| SOS-mediated signalling | 185 | 19 (10.3%) | 0.00551 | 0.0456 | Reactome | 2.26 |
| GRB2 events in EGFR signaling | 185 | 19 (10.3%) | 0.00551 | 0.0456 | Reactome | 2.26 |
| SHC1 events in ERBB2 signaling | 185 | 19 (10.3%) | 0.00551 | 0.0456 | Reactome | 2.26 |
| SHC1 events in ERBB4 signaling | 185 | 19 (10.3%) | 0.00551 | 0.0456 | Reactome | 2.26 |
| GRB2 events in ERBB2 signaling | 185 | 19 (10.3%) | 0.00551 | 0.0456 | Reactome | 2.26 |
| Receptor-ligand binding initiates the second proteolytic cleavage of Notch receptor | 14 | 4 (28.6%) | 0.00565 | 0.0463 | Reactome | 2.25 |
| FRS-mediated FGFR2 signaling | 186 | 19 (10.3%) | 0.00584 | 0.0469 | Reactome | 2.23 |
| FRS-mediated FGFR1 signaling | 186 | 19 (10.3%) | 0.00584 | 0.0469 | Reactome | 2.23 |
| FRS-mediated FGFR3 signaling | 186 | 19 (10.3%) | 0.00584 | 0.0469 | Reactome | 2.23 |
| FRS-mediated FGFR4 signaling | 186 | 19 (10.3%) | 0.00584 | 0.0469 | Reactome | 2.23 |
| Interleukin-2 signaling | 202 | 20 (10.0%) | 0.00674 | 0.0533 | Reactome | 2.17 |
| Signalling to p38 via RIT and RIN | 189 | 19 (10.1%) | 0.00693 | 0.0536 | Reactome | 2.16 |
| ARMS-mediated activation | 189 | 19 (10.1%) | 0.00693 | 0.0536 | Reactome | 2.16 |
| Adaptive Immune System | 612 | 47 (7.9%) | 0.00702 | 0.0538 | Reactome | 2.15 |
| RAF-independent MAPK1/3 activation | 23 | 5 (21.7%) | 0.00703 | 0.0538 | Reactome | 2.15 |
| Retinoid metabolism and transport | 42 | 7 (16.7%) | 0.00705 | 0.0538 | Reactome | 2.15 |
| Frs2-mediated activation | 190 | 19 (10.1%) | 0.00732 | 0.0552 | Reactome | 2.14 |
| RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways | 53 | 8 (15.1%) | 0.00753 | 0.0556 | Reactome | 2.12 |
| Prolonged ERK activation events | 192 | 19 (9.9%) | 0.00817 | 0.06 | Reactome | 2.09 |
| GPVI-mediated activation cascade | 54 | 8 (14.8%) | 0.00843 | 0.0611 | Reactome | 2.07 |
| Interconversion of polyamines | 3 | 2 (66.7%) | 0.00858 | 0.0616 | Reactome | 2.07 |
| Hydroxycarboxylic acid-binding receptors | 3 | 2 (66.7%) | 0.00858 | 0.0616 | Reactome | 2.07 |
| Signalling to RAS | 194 | 19 (9.8%) | 0.00909 | 0.065 | Reactome | 2.04 |
| Integrin cell surface interactions | 67 | 9 (13.6%) | 0.0092 | 0.0655 | Reactome | 2.04 |
| Interleukin receptor SHC signaling | 195 | 19 (9.8%) | 0.00958 | 0.0679 | Reactome | 2.02 |

**Table S5.** Candidaemia-associated SNPs and variants with $r^2$>=0.8. SNPs rs12491812 and rs1802141 were in strong linkage disequilibrium (LD) with synonymous variants located in *CISH* and *SNPS1* genes respectively (source: Haploreg http://archive.broadinstitute.org/mammals/haploreg/haploreg.php).

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 18989152 | 0.9 | 0.95 | rs57825564 | A | G | 0.11 | 32kb 5' of *IFFO2* | Intergenic |
| 1 | 18989323 | 0.9 | 0.95 | rs59569385 | A | G | 0.11 | 33kb 5' of *IFFO2* | Intergenic |
| 1 | 18990505 | 0.91 | 0.96 | rs59405037 | T | C | 0.11 | 34kb 5' of *IFFO2* | Intergenic |
| 1 | 18990540 | 0.91 | 0.96 | rs58552530 | G | T | 0.11 | 34kb 5' of *IFFO2* | Intergenic |
| 1 | 18990607 | 0.9 | 0.96 | rs57735998 | A | G | 0.12 | 34kb 5' of *IFFO2* | Intergenic |
| 1 | 18990844 | 0.91 | 0.96 | rs28789190 | G | A | 0.11 | 34kb 5' of *IFFO2* | Intergenic |
| 1 | 18991035 | 0.91 | 0.96 | rs6669987 | G | A | 0.11 | 34kb 5' of *IFFO2* | Intergenic |
| 1 | 18991493 | 0.91 | 0.96 | rs7527267 | G | T | 0.11 | 35kb 5' of *IFFO2* | Intergenic |
| 1 | 18991593 | 0.91 | 0.96 | rs7540747 | A | G | 0.11 | 35kb 5' of *IFFO2* | Intergenic |
| 1 | 18991819 | 0.89 | 0.96 | rs6682824 | C | G | 0.12 | 35kb 5' of *IFFO2* | Intergenic |
| 1 | 18992603 | 0.92 | 0.96 | rs7539506 | C | A | 0.11 | 36kb 5' of *IFFO2* | Intergenic |
| 1 | 18992627 | 0.92 | 0.96 | rs7543956 | A | T | 0.11 | 36kb 5' of *IFFO2* | Intergenic |
| 1 | 18993125 | 0.92 | 0.96 | rs6694309 | T | C | 0.11 | 36kb 5' of *IFFO2* | Intergenic |
| 1 | 18993715 | 0.92 | 0.96 | rs6697292 | T | G | 0.11 | 37kb 5' of *IFFO2* | Intergenic |
| 1 | 18993825 | 0.92 | 0.96 | rs6689626 | C | A | 0.11 | 37kb 5' of *IFFO2* | Intergenic |
| 1 | 18993937 | 0.92 | 0.96 | rs6694529 | A | G | 0.11 | 37kb 5' of *IFFO2* | Intergenic |
| 1 | 18994050 | 0.92 | 0.96 | rs6689820 | C | A | 0.11 | 37kb 5' of *IFFO2* | Intergenic |
| 1 | 18994344 | 0.92 | 0.96 | rs28623160 | T | C | 0.11 | 38kb 5' of *IFFO2* | Intergenic |
| 1 | 18994421 | 0.92 | 0.96 | rs28575256 | C | T | 0.11 | 38kb 5' of *IFFO2* | Intergenic |
| 1 | 18994571 | 0.92 | 0.96 | rs28651004 | A | G | 0.11 | 38kb 5' of *IFFO2* | Intergenic |
| 1 | 18995310 | 0.95 | 0.97 | rs28369012 | A | C | 0.11 | 39kb 5' of *IFFO2* | Intergenic |
| 1 | 18995575 | 0.95 | 0.97 | rs56153612 | C | T | 0.11 | 39kb 5' of *IFFO2* | Intergenic |
| 1 | 18995992 | 0.95 | 0.97 | rs4611046 | A | G | 0.11 | 39kb 5' of *IFFO2* | Intergenic |
| 1 | 18996055 | 0.92 | 0.97 | rs1570861 | G | A | 0.12 | 39kb 5' of *IFFO2* | Intergenic |
| 1 | 18996157 | 0.94 | 0.97 | rs1570860 | T | C | 0.11 | 39kb 5' of *IFFO2* | Intergenic |
| 1 | 18996556 | 0.94 | 0.97 | rs6690097 | G | C | 0.11 | 40kb 5' of *IFFO2* | Intergenic |
| 1 | 18996839 | 0.94 | 0.97 | rs6659084 | T | C | 0.11 | 40kb 5' of *IFFO2* | Intergenic |
| 1 | 18996887 | 0.94 | 0.97 | rs6690430 | G | T | 0.11 | 40kb 5' of *IFFO2* | Intergenic |
| 1 | 18997108 | 0.9 | 0.95 | rs1316768 | G | A | 0.11 | 40kb 5' of *IFFO2* | Intergenic |
| 1 | 18997598 | 0.9 | 0.96 | rs6703129 | C | T | 0.12 | 41kb 5' of *IFFO2* | Intergenic |
| 1 | 18997744 | 0.94 | 0.97 | rs28690197 | C | G | 0.11 | 41kb 5' of *IFFO2* | Intergenic |
| 1 | 18997899 | 0.94 | 0.97 | rs28417827 | G | A | 0.11 | 41kb 5' of *IFFO2* | Intergenic |
| 1 | 18997917 | 0.94 | 0.97 | rs28686865 | G | A | 0.11 | 41kb 5' of *IFFO2* | Intergenic |
| 1 | 18998168 | 0.94 | 0.97 | rs28431045 | T | C | 0.11 | 41kb 5' of *IFFO2* | Intergenic |
| 1 | 18998191 | 0.94 | 0.97 | rs28429352 | C | A | 0.11 | 42kb 5' of *IFFO2* | Intergenic |
| 1 | 18998227 | 0.91 | 0.96 | rs28366691 | G | A | 0.11 | 42kb 5' of *IFFO2* | Intergenic |
| 1 | 18998229 | 0.94 | 0.97 | rs28628464 | C | T | 0.11 | 42kb 5' of *IFFO2* | Intergenic |
| 1 | 18998590 | 0.94 | 0.97 | rs28715564 | A | G | 0.11 | 42kb 5' of *IFFO2* | Intergenic |
| 1 | 18999277 | 0.97 | 1 | rs6700076 | G | C | 0.11 | 43kb 5' of *IFFO2* | Intergenic |
| 1 | 18999422 | 0.97 | 0.99 | rs6666106 | A | G | 0.11 | 43kb 5' of *IFFO2* | Intergenic |
| 1 | 18999740 | 0.95 | 0.99 | rs6669421 | T | A | 0.12 | 43kb 5' of *IFFO2* | Intergenic |
| 1 | 18999819 | 0.97 | 0.99 | rs6700619 | G | C | 0.11 | 43kb 5' of *IFFO2* | Intergenic |
| 1 | 19000277 | 1 | 1 | rs28398771 | A | G | 0.11 | 44kb 5' of *IFFO2* | Intergenic |
| 1 | 19000319 | 1 | 1 | rs6699706 | A | C | 0.11 | 44kb 5' of *IFFO2* | Intergenic |
| 1 | 19000888 | 1 | 1 | rs6700354 | A | G | 0.11 | 44kb 5' of *IFFO2* | Intergenic |
| 1 | 19001112 | 0.81 | 1 | rs6703643 | T | G | 0.09 | 44kb 5' of *IFFO2* | Intergenic |
| 1 | 19001271 | 1 | 1 | rs72951502 | C | T | 0.11 | 45kb 5' of *IFFO2* | Intergenic |
| 1 | 19001288 | 1 | 1 | rs28382321 | G | A | 0.11 | 45kb 5' of *IFFO2* | Intergenic |
| 1 | 19001345 | 1 | 1 | rs28480631 | T | C | 0.11 | 45kb 5' of *IFFO2* | Intergenic |
| 1 | 19001583 | 1 | 1 | rs28806527 | T | C | 0.11 | 45kb 5' of *IFFO2* | Intergenic |
| 1 | 19001748 | 1 | 1 | rs28840430 | T | C | 0.11 | 45kb 5' of *IFFO2* | Intergenic |
| 1 | 19001964 | 1 | 1 | rs28505299 | G | T | 0.11 | 45kb 5' of *IFFO2* | Intergenic |
| 1 | 19001966 | 1 | 1 | rs28633203 | A | G | 0.11 | 45kb 5' of *IFFO2* | Intergenic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 19002512 | 1 | 1 | rs6701963 | C | G | 0.11 | 46kb 5' of *IFFO2* | Intergenic |
| 1 | 19003051 | 1 | 1 | rs6702530 | C | T | 0.11 | 46kb 5' of *IFFO2* | Intergenic |
| 1 | 19003218 | 1 | 1 | rs28505711 | G | A | 0.11 | 47kb 5' of *IFFO2* | Intergenic |
| 1 | 19004562 | 0.97 | 1 | rs2012808 | A | G | 0.12 | 48kb 5' of *IFFO2* | Intergenic |
| 1 | 19004650 | 1 | 1 | rs6699361 | G | T | 0.11 | 48kb 5' of *IFFO2* | Intergenic |
| 1 | 19004914 | 1 | 1 | rs1078895 | T | C | 0.11 | 48kb 5' of *IFFO2* | Intergenic |
| 1 | 19005103 | 1 | 1 | rs1078893 | T | A | 0.11 | 48kb 5' of *IFFO2* | Intergenic |
| 1 | 19005397 | 1 | 1 | rs6426622 | T | C | 0.11 | 49kb 5' of *IFFO2* | Intergenic |
| 1 | 19005582 | 0.99 | 1 | rs7523455 | A | G | 0.11 | 49kb 5' of *IFFO2* | Intergenic |
| 1 | 19006097 | 0.96 | 0.99 | rs16862451 | C | T | 0.11 | 49kb 5' of *IFFO2* | Intergenic |
| 1 | 19006231 | 0.96 | 0.99 | rs28675749 | C | T | 0.11 | 50kb 5' of *IFFO2* | Intergenic |
| 1 | 19006280 | 0.96 | 0.99 | rs16862452 | T | C | 0.11 | 50kb 5' of *IFFO2* | Intergenic |
| **Locus 2. Query SNP: rs11102637 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 1 | 113405301 | 0.8 | 0.9 | rs1418607 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113405986 | 0.8 | 0.9 | rs2027537 | C | G,T | 0.18 | *MAGI3* | Intronic |
| 1 | 113406706 | 0.8 | 0.9 | rs12021783 | A | C | 0.18 | *MAGI3* | Intronic |
| 1 | 113408485 | 0.8 | 0.9 | rs12023641 | A | C | 0.18 | *MAGI3* | Intronic |
| 1 | 113410914 | 0.8 | 0.9 | rs55737954 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113424319 | 0.82 | 0.91 | rs10458458 | G | C | 0.18 | *MAGI3* | Intronic |
| 1 | 113424783 | 0.83 | 0.91 | rs10458459 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113425052 | 0.81 | 0.91 | rs12029305 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113428152 | 0.83 | 0.91 | rs12032395 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113429925 | 0.82 | 0.91 | rs72687924 | G | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113431641 | 0.82 | 0.91 | rs12030900 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113431937 | 0.82 | 0.91 | rs12034178 | G | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113436043 | 0.82 | 0.91 | rs35818543 | CAT | C | 0.18 | *MAGI3* | Intronic |
| 1 | 113437599 | 0.82 | 0.91 | rs5777159 | TA | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113440005 | 1 | 1 | rs12039669 | T | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113441032 | 1 | 1 | rs12022213 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113441121 | 1 | 1 | rs12039944 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113441435 | 1 | 1 | rs1936928 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113441608 | 1 | 1 | rs1936927 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113442834 | 0.99 | 1 | rs1573996 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113446628 | 1 | 1 | rs11102636 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113447427 | 0.99 | 1 | rs12043827 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113449170 | 1 | 1 | rs11102637 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113449853 | 1 | 1 | rs12025806 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113454244 | 0.99 | 1 | rs12045559 | G | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113454850 | 1 | 1 | rs12046450 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113455064 | 1 | 1 | rs11102638 | T | C | 0.18 | *MAGI3* | Intronic |
| 1 | 113464525 | 1 | 1 | rs11102640 | G | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113465662 | 1 | 1 | rs11102641 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113468315 | 1 | 1 | rs72687938 | T | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113468684 | 1 | 1 | rs72687939 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113473636 | 1 | 1 | rs78995277 | T | C | 0.18 | *MAGI3* | Intronic |
| | | 0.96 | 0.98 | rs201975600 | TTC | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113476614 | 1 | 1 | rs148488077 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113477369 | 1 | 1 | rs7413018 | T | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113480445 | 1 | 1 | rs74465571 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113480717 | 1 | 1 | rs72687944 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113481726 | 1 | 1 | rs12038473 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113482047 | 1 | 1 | rs11102644 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113484093 | 1 | 1 | rs72687947 | G | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113484144 | 1 | 1 | rs72687948 | A | C | 0.18 | *MAGI3* | Intronic |
| 1 | 113484858 | 1 | 1 | rs72687949 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113488304 | 1 | 1 | rs12044485 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113489889 | 1 | 1 | rs12039226 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113492466 | 1 | 1 | rs12033666 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113493543 | 1 | 1 | rs1936926 | G | A | 0.18 | *MAGI3* | Intronic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 113507469 | 0.95 | 0.97 | rs12046451 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113513195 | 0.91 | 0.95 | rs74988614 | T | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113513394 | 0.82 | 0.95 | rs12046434 | G | A | 0.16 | *MAGI3* | Intronic |
| 1 | 113515327 | 0.91 | 0.95 | rs12045148 | A | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113515658 | 0.89 | 0.95 | rs4838988 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113519247 | 0.89 | 0.95 | rs12024139 | G | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113520568 | 0.87 | 0.95 | rs11102645 | T | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113520901 | 0.89 | 0.95 | rs12022849 | A | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113522168 | 0.89 | 0.95 | rs2884705 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113522276 | 0.88 | 0.95 | rs2359413 | T | G | 0.18 | *MAGI3* | Intronic |
| 1 | 113522673 | 0.89 | 0.95 | rs72687960 | T | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113522718 | 0.89 | 0.95 | rs72687961 | C | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113525957 | 0.89 | 0.95 | rs2051083 | G | T | 0.18 | *MAGI3* | Intronic |
| 1 | 113526463 | 0.89 | 0.95 | rs12026682 | G | A | 0.18 | *MAGI3* | Intronic |
| 1 | 113528995 | 0.85 | 0.94 | rs4839324 | G | A | 0.17 | *MAGI3* | Intronic |
| **Locus 3. Query SNP: rs3766122 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 1 | 169555933 | 0.86 | 0.97 | rs9332581 | G | T | 0.05 | *F5* | Intronic |
| 1 | 169556011 | 0.86 | 0.97 | rs9332580 | A | G | 0.05 | *F5* | Intronic |
| 1 | 169566680 | 0.91 | 0.98 | rs12024736 | A | G | 0.06 | *F5* | Intronic |
| 1 | 169591019 | 0.95 | 1 | rs3917843 | C | T | 0.06 | *SELP* | Intronic |
| 1 | 169597589 | 1 | 1 | rs3766122 | T | C | 0.06 | *SELP* | Intronic |
| **Locus 4. Query SNP: rs296537 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 1 | 200919286 | 1 | -1 | rs78154166 | G | A | 0.01 | 3.6kb 3' of *C1orf106* | Intergenic |
| 1 | 200920121 | 1 | -1 | rs139023910 | A | T | 0.01 | 4.4kb 3' of *C1orf106* | Intergenic |
| 1 | 200921348 | 1 | -1 | rs59233488 | C | G | 0.01 | 5.6kb 3' of *C1orf106* | Intergenic |
| 1 | 200925926 | 1 | -1 | rs113661670 | A | G | 0.01 | 7.4kb 3' of *U6* | Intergenic |
| 1 | 200931047 | 1 | -1 | rs143611860 | G | C | 0.01 | 2.3kb 3' of *U6* | Intronic |
| 1 | 200935921 | 1 | -1 | rs13374187 | A | G | 0.01 | 2.5kb 5' of *U6* | Intronic |
| 1 | 200936507 | 1 | -1 | rs113689739 | A | T | 0.01 | 3.1kb 5' of *U6* | Intronic |
| 1 | 200936759 | 1 | -1 | rs7525822 | A | G | 0.01 | 3.3kb 5' of *U6* | Intronic |
| 1 | 200937554 | 1 | -1 | rs146936361 | G | A | 0.01 | 4.1kb 5' of *U6* | Intronic |
| 1 | 200942108 | 1 | -1 | rs145480224 | G | C | 0.01 | 8.7kb 5' of *U6* | Intronic |
| 1 | 200942917 | 1 | -1 | rs10920073 | A | G | 0.01 | 9.5kb 5' of *U6* | Intronic |
| 1 | 200944808 | 1 | -1 | rs10920075 | T | A | 0.01 | 11kb 5' of *U6* | Intronic |
| 1 | 200948002 | 1 | 1 | rs296537 | A | G | 0.99 | 15kb 5' of *U6* | Intronic |
| 1 | 200948169 | 1 | -1 | rs10920077 | G | A | 0.01 | 15kb 5' of *U6* | Intergenic |
| 1 | 200949719 | 1 | -1 | rs10920078 | C | T | 0.01 | 16kb 5' of *U6* | Intronic |
| 1 | 200950832 | 1 | -1 | rs80051606 | A | C | 0.01 | 17kb 5' of *U6* | Intronic |
| 1 | 200952049 | 1 | -1 | rs111635100 | C | T | 0.01 | 17kb 3' of *KIF21B* | Intronic |
| 1 | 200952267 | 1 | -1 | rs28445895 | C | A | 0.01 | 17kb 3' of *KIF21B* | Intronic |
| 1 | 200952668 | 1 | -1 | rs12079204 | A | G | 0.01 | 17kb 3' of *KIF21B* | Intronic |
| 1 | 200952817 | 1 | -1 | rs12074541 | T | C | 0.01 | 17kb 3' of *KIF21B* | Intronic |
| 1 | 200958620 | 0.83 | -1 | rs16847413 | G | A | 0.01 | 11kb 3' of *KIF21B* | Intronic |
| 1 | 200959658 | 0.83 | -1 | rs7554653 | T | C | 0.01 | 9.7kb 3' of *KIF21B* | Intronic |
| 1 | 201022257 | 1 | -1 | rs139357431 | G | A | 0.01 | *KIF21B* | Intronic |
| **Locus 5. Query SNP: rs6748999 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 2 | 127447412 | 0.81 | -0.9 | rs6745666 | C | G | 0.96 | *IWS1* | Intergenic |
| 2 | 127447787 | 0.81 | -0.9 | rs6716856 | G | T | 0.96 | *IWS1* | Intergenic |
| 2 | 127455012 | 0.81 | 0.9 | rs777567 | G | T | 0.04 | *IWS1* | Intergenic |
| 2 | 127455583 | 0.81 | 0.9 | rs77511630 | C | G | 0.04 | *IWS1* | Intergenic |
| 2 | 127469132 | 1 | 1 | rs6748999 | A | G | 0.04 | *IWS1* | Intergenic |
| 2 | 127506177 | 0.85 | 0.93 | rs140923032 | G | A | 0.04 | *IWS1* | Intronic |
| 2 | 127507153 | 0.85 | 0.93 | rs114358254 | T | C | 0.04 | *IWS1* | Intronic |
| 2 | 127512132 | 0.82 | 0.93 | rs75704656 | G | A | 0.04 | *IWS1* | Intronic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| **Locus 6. Query SNP: rs12491812 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 3 | 50497204 | 0.82 | 1 | rs9814874 | G | A | 0.01 | *CACNA2D2* | Intronic |
| 3 | 50501788 | 0.82 | 1 | rs1107312 | A | G | 0.01 | *CACNA2D2* | Intronic |
| 3 | 50505295 | 0.82 | 1 | rs3806706 | G | C | 0.01 | 1.1kb 5' of *CACNA2D2* | Intergenic |
| 3 | 50518502 | 1 | 1 | rs17050991 | A | C | 0.01 | 14kb 5' of *CACNA2D2* | Intergenic |
| 3 | 50519150 | 1 | 1 | rs12491812 | C | T | 0.01 | 15kb 5' of *CACNA2D2* | intergenic |
| 3 | 50521386 | 1 | 1 | rs61574029 | T | G | 0.01 | 17kb 5' of *CACNA2D2* | intergenic |
| 3 | 50522770 | 1 | 1 | rs61613532 | T | C | 0.01 | 19kb 5' of *CACNA2D2* | intergenic |
| 3 | 50523067 | 1 | 1 | rs61193866 | G | A | 0.01 | 19kb 5' of *CACNA2D2* | intergenic |
| 3 | 50523127 | 0.82 | 1 | rs72938758 | A | G | 0.01 | 19kb 5' of *CACNA2D2* | intergenic |
| 3 | 50525135 | 1 | 1 | rs12487817 | T | C | 0.01 | 21kb 5' of *CACNA2D2* | intergenic |
| 3 | 50526739 | 1 | 1 | rs12486143 | G | T | 0.01 | 22kb 5' of *CACNA2D2* | intergenic |
| 3 | 50529066 | 1 | 1 | rs4688713 | A | G | 0.01 | 25kb 5' of *CACNA2D2* | intergenic |
| 3 | 50529104 | 1 | -1 | rs6766638 | G | T | 0.99 | 25kb 5' of *CACNA2D2* | intergenic |
| 3 | 50530757 | 1 | 1 | rs73835531 | T | C | 0.01 | 27kb 5' of *CACNA2D2* | intergenic |
| 3 | 50532344 | 1 | 1 | rs17050995 | A | G | 0.01 | 26kb 3' of *C3orf18* | intergenic |
| 3 | 50540850 | 0.89 | 1 | rs17051000 | A | T | 0.01 | 17kb 3' of *C3orf18* | intergenic |
| 3 | 50544526 | 0.89 | 1 | rs12494094 | A | C | 0.01 | 14kb 3' of *C3orf18* | intergenic |
| 3 | 50550745 | 0.9 | 1 | rs56170012 | A | G | 0.01 | 7.3kb 3' of *C3orf18* | intergenic |
| 3 | 50551935 | 0.89 | 1 | rs60552073 | C | T | 0.01 | 6.1kb 3' of *C3orf18* | intergenic |
| 3 | 50553197 | 0.89 | 1 | rs4688710 | G | A | 0.01 | 4.8kb 3' of *C3orf18* | intergenic |
| 3 | 50571017 | 0.89 | 1 | rs2227293 | G | A | 0.01 | *C3orf18* | 5'-UTR |
| 3 | 50582612 | 0.89 | 1 | rs2285090 | T | C | 0.01 | *HEMK1* | 3'-UTR |
| 3 | 50584763 | 0.89 | 1 | rs72341372 | CAT | C | 0.01 | *HEMK1* | 3'-UTR |
| 3 | 50586144 | 0.89 | 1 | rs3749265 | T | A | 0.01 | 1.2kb 3' of *HEMK1* | intergenic |
| 3 | 50586245 | 0.89 | 1 | rs1894459 | A | G | 0.01 | 1.3kb 3' of *HEMK1* | intergenic |
| 3 | 50589619 | 0.89 | 1 | rs79192940 | G | A | 0.01 | 4.7kb 3' of *HEMK1* | intergenic |
| 3 | 50592228 | 0.89 | 1 | rs12494326 | C | G | 0.01 | 7.3kb 3' of *HEMK1* | intergenic |
| 3 | 50596404 | 0.89 | 1 | rs61316441 | C | T | 0.01 | 10kb 3' of *CISH* | intergenic |
| 3 | 50596595 | 0.89 | 1 | rs4688707 | C | T | 0.01 | 9.9kb 3' of *CISH* | intergenic |
| 3 | 50607727 | 0.89 | 1 | rs2239753 | A | G | 0.01 | *CISH* | synonymous |
| 3 | 50607982 | 0.89 | 1 | rs2239752 | G | A | 0.01 | *CISH* | synonymous |
| **Locus 7. Query SNP: rs16891982 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 5 | 33951588 | 1 | 1 | rs16891982 | C | G | 0.97 | *SLC45A2* | intergenic |
| **Locus 8. Query SNP: rs11760176 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 6 | 29980813 | 0.84 | 0.97 | rs139093747 | GA | G | 0.04 | 2.4kb 3' of *HCG9* | intergenic |
| 6 | 29982544 | 0.84 | 0.97 | rs3734830 | G | A | 0.04 | 4.1kb 3' of *HCG9* | intergenic |
| 6 | 29982640 | 0.84 | 0.97 | rs3734831 | T | C | 0.04 | 4.2kb 3' of *HCG9* | intergenic |
| 6 | 29982800 | 0.84 | 0.97 | rs3734833 | A | G | 0.04 | 4.4kb 3' of *HCG9* | intergenic |
| 6 | 29982804 | 0.86 | 0.96 | rs3734834 | A | G | 0.04 | 4.4kb 3' of *HCG9* | intergenic |
| 6 | 29983062 | 0.84 | 0.97 | rs73725904 | G | C | 0.04 | 4.7kb 3' of *HCG9* | intergenic |
| 6 | 29983202 | 0.84 | 0.97 | rs73428121 | A | C | 0.04 | 4.8kb 3' of *HCG9* | intergenic |
| 6 | 29983272 | 0.84 | 0.97 | rs73428122 | T | G | 0.04 | 4.9kb 3' of *HCG9* | intergenic |
| 6 | 29983284 | 0.81 | 0.93 | rs114832360 | C | T | 0.04 | 4.9kb 3' of *HCG9* | intergenic |
| 6 | 29983307 | 0.81 | 0.93 | rs73428125 | T | A | 0.04 | 4.9kb 3' of *HCG9* | intergenic |
| 6 | 29983427 | 0.84 | 0.97 | rs57035982 | G | A | 0.04 | 5kb 3' of *HCG9* | intergenic |
| 6 | 29983496 | 0.84 | 0.97 | rs56015532 | A | G | 0.04 | 5.1kb 3' of *HCG9* | intergenic |
| 6 | 29983506 | 0.84 | 0.97 | rs57080532 | A | C | 0.04 | 5.1kb 3' of *HCG9* | intergenic |
| 6 | 29983517 | 0.84 | 0.97 | rs57680173 | G | T | 0.04 | 5.1kb 3' of *HCG9* | intergenic |
| 6 | 29983834 | 0.81 | 0.93 | rs55760574 | A | T | 0.04 | 5.4kb 3' of *HCG9* | intergenic |
| 6 | 29983842 | 0.81 | 0.93 | rs60587228 | G | A | 0.04 | 5.4kb 3' of *HCG9* | intergenic |
| 6 | 29983902 | 0.81 | 0.93 | rs56385845 | G | A | 0.04 | 5.5kb 3' of *HCG9* | intergenic |
| 6 | 29984114 | 0.84 | 0.97 | rs58161185 | A | G | 0.04 | 5.7kb 3' of *HCG9* | intergenic |
| 6 | 29984287 | 0.87 | 0.97 | rs201313489 | CT | C | 0.04 | 5.9kb 3' of *HCG9* | intergenic |
| 6 | 29984292 | 0.87 | 0.97 | rs35361490 | G | A | 0.04 | 5.9kb 3' of *HCG9* | intergenic |
| 6 | 29984325 | 0.87 | 0.97 | rs34513540 | G | T | 0.04 | 5.9kb 3' of *HCG9* | intergenic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 29984329 | 0.84 | 0.93 | rs114426132 | C | T | 0.04 | 5.9kb 3' of *HCG9* | intergenic |
| 6 | 29984330 | 0.84 | 0.93 | rs115313680 | C | G | 0.04 | 5.9kb 3' of *HCG9* | intergenic |
| 6 | 29984370 | 0.84 | 0.97 | rs11759891 | A | T | 0.04 | 6kb 3' of *HCG9* | intergenic |
| 6 | 29984385 | 0.84 | 0.97 | rs11757750 | T | C | 0.04 | 6kb 3' of *HCG9* | intergenic |
| 6 | 29984538 | 0.81 | 0.93 | rs11755337 | C | G | 0.04 | 6.1kb 3' of *HCG9* | intergenic |
| 6 | 29984548 | 0.81 | 0.93 | rs11757792 | T | C | 0.04 | 6.1kb 3' of *HCG9* | intergenic |
| 6 | 29984582 | 0.84 | 0.97 | rs11755374 | G | A | 0.04 | 6.2kb 3' of *HCG9* | intergenic |
| 6 | 29984642 | 0.84 | 0.97 | rs61704409 | C | T | 0.04 | 6.2kb 3' of *HCG9* | intergenic |
| 6 | 29984705 | 0.84 | 0.97 | rs35144883 | C | T | 0.04 | 6.3kb 3' of *HCG9* | intergenic |
| 6 | 29984776 | 0.84 | 0.97 | rs58867136 | A | G | 0.04 | 6.4kb 3' of *HCG9* | intergenic |
| 6 | 29984927 | 0.84 | 0.97 | rs17186930 | C | T | 0.04 | 6.5kb 3' of *HCG9* | intergenic |
| 6 | 29985037 | 0.84 | 0.97 | rs17186937 | T | C | 0.04 | 6.6kb 3' of *HCG9* | intergenic |
| 6 | 29985143 | 0.84 | 0.97 | rs17186944 | C | T | 0.04 | 6.7kb 3' of *HCG9* | intergenic |
| 6 | 29985154 | 0.84 | 0.97 | rs11755966 | G | A | 0.04 | 6.7kb 3' of *HCG9* | intergenic |
| 6 | 29985258 | 0.84 | 0.97 | rs11755984 | G | A | 0.04 | 6.9kb 3' of *HCG9* | intergenic |
| 6 | 29985322 | 0.84 | 0.97 | rs11755961 | C | T | 0.04 | 6.9kb 3' of *HCG9* | intergenic |
| 6 | 29985493 | 0.84 | 0.97 | rs11755992 | C | G | 0.04 | 7.1kb 3' of *HCG9* | intergenic |
| 6 | 29985534 | 0.84 | 0.97 | rs11751331 | A | T | 0.04 | 7.1kb 3' of *HCG9* | intergenic |
| 6 | 29985651 | 0.84 | 0.97 | rs11758493 | T | G | 0.04 | 7.2kb 3' of *HCG9* | intergenic |
| 6 | 29985849 | 0.84 | 0.97 | rs73725909 | C | T | 0.04 | 7.4kb 3' of *HCG9* | intergenic |
| 6 | 29985851 | 0.81 | 0.93 | rs73725910 | G | A | 0.04 | 7.4kb 3' of *HCG9* | intergenic |
| 6 | 29985941 | 0.81 | 0.93 | rs73725912 | G | A | 0.04 | 7.5kb 3' of *HCG9* | intergenic |
| 6 | 29986026 | 0.84 | 0.97 | rs73428152 | C | A | 0.04 | 7.6kb 3' of *HCG9* | intergenic |
| 6 | 29986208 | 0.84 | 0.97 | rs5875226 | C | CA | 0.04 | 7.8kb 3' of *HCG9* | intergenic |
| 6 | 29986211 | 0.81 | 0.93 | rs35685401 | C | T | 0.04 | 7.8kb 3' of *HCG9* | intergenic |
| 6 | 29986334 | 0.84 | 0.97 | rs12662947 | T | C | 0.04 | 7.9kb 3' of *HCG9* | intergenic |
| 6 | 29986422 | 0.84 | 0.97 | rs12661411 | C | T | 0.04 | 8kb 3' of *HCG9* | intergenic |
| 6 | 29986503 | 0.84 | 0.97 | rs12662967 | T | C | 0.04 | 8.1kb 3' of *HCG9* | intergenic |
| 6 | 29986679 | 0.81 | 0.93 | rs73410503 | G | A | 0.04 | 8.3kb 3' of *HCG9* | intergenic |
| 6 | 29986840 | 0.84 | 0.97 | rs34894339 | A | G | 0.04 | 8.4kb 3' of *HCG9* | intergenic |
| 6 | 29986847 | 0.84 | 0.97 | rs73725917 | A | C,T | 0.04 | 8.4kb 3' of *HCG9* | intergenic |
| 6 | 29987074 | 0.84 | 0.97 | rs17186979 | A | G | 0.04 | 8.7kb 3' of *HCG9* | intergenic |
| 6 | 29987280 | 0.84 | 0.97 | rs17187021 | T | C | 0.04 | 8.9kb 3' of *HCG9* | intergenic |
| 6 | 29987310 | 0.81 | 0.93 | rs17187035 | T | C | 0.04 | 8.9kb 3' of *HCG9* | intergenic |
| 6 | 29987445 | 0.84 | 0.97 | rs17187077 | A | C | 0.04 | 9kb 3' of *HCG9* | intergenic |
| 6 | 29987611 | 0.81 | 0.93 | rs17187141 | T | C | 0.04 | 9.2kb 3' of *HCG9* | intergenic |
| 6 | 29987762 | 0.84 | 0.97 | rs17180794 | T | C | 0.04 | 9.4kb 3' of *HCG9* | intergenic |
| 6 | 29987833 | 0.8 | 0.9 | rs73410515 | G | A | 0.04 | 9.4kb 3' of *HCG9* | intergenic |
| 6 | 29987997 | 0.84 | 0.97 | rs73410518 | T | C,G | 0.04 | 9.6kb 3' of *HCG9* | intergenic |
| 6 | 29988243 | 0.84 | 0.97 | rs76517404 | T | C | 0.04 | 9.8kb 3' of *HCG9* | intergenic |
| 6 | 29988248 | 0.81 | 0.93 | rs201767378 | AG | A | 0.04 | 9.8kb 3' of *HCG9* | intergenic |
| 6 | 29988272 | 0.84 | 0.97 | rs73410524 | G | A | 0.04 | 9.9kb 3' of *HCG9* | intergenic |
| 6 | 29988401 | 0.84 | 0.97 | rs73410525 | C | A | 0.04 | 10kb 3' of *HCG9* | intergenic |
| 6 | 29988516 | 0.84 | 0.97 | rs112793716 | C | T | 0.04 | 10kb 3' of *HCG9* | intergenic |
| 6 | 29988552 | 0.84 | 0.97 | rs73410528 | G | A | 0.04 | 10kb 3' of *HCG9* | intergenic |
| 6 | 29988698 | 0.81 | 0.93 | rs147065659 | TGAG | T | 0.04 | 10kb 3' of *HCG9* | intergenic |
| 6 | 29988943 | 0.84 | 0.97 | rs12662336 | G | A | 0.04 | 11kb 3' of *HCG9* | intergenic |
| 6 | 29989300 | 0.84 | 0.97 | rs12662611 | C | T | 0.04 | 11kb 3' of *HCG9* | intergenic |
| 6 | 29989508 | 0.84 | 0.97 | rs12664166 | T | G | 0.04 | 11kb 3' of *HCG9* | intergenic |
| 6 | 29993802 | 0.84 | 0.97 | rs17187224 | C | T | 0.04 | 7.2kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29994016 | 0.84 | 0.97 | rs17180864 | C | T | 0.04 | 7kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29994066 | 0.84 | 0.97 | rs17187252 | G | A | 0.04 | 6.9kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29995744 | 0.84 | 0.97 | rs6911583 | T | C | 0.04 | 5.3kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29995947 | 0.84 | 0.97 | rs58858573 | C | A | 0.04 | 5.1kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29996620 | 0.84 | 0.97 | rs10947040 | G | A | 0.04 | 4.4kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29997052 | 0.84 | 0.97 | rs12110437 | A | T | 0.04 | 4kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29998038 | 0.84 | 0.97 | rs72502507 | A | C | 0.04 | 3kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29998247 | 0.81 | 0.93 | rs12111210 | G | A | 0.04 | 2.8kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29998893 | 0.81 | 0.93 | rs6917877 | A | G | 0.04 | 2.1kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 29998897 | 0.84 | 0.97 | rs6940956 | T | A | 0.04 | 2.1kb 3' of *ZNRD1-AS1* | intergenic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 29999555 | 0.81 | 0.93 | rs17187349 | T | C | 0.04 | 1.5kb 3' of *ZNRD1-AS1* | intergenic |
| 6 | 30000403 | 0.81 | 0.93 | rs12664967 | C | G | 0.04 | 607bp 3' of *ZNRD1-AS1* | intergenic |
| 6 | 30000678 | 0.87 | 0.97 | rs10947042 | C | T | 0.04 | 332bp 3' of *ZNRD1-AS1* | intergenic |
| 6 | 30000900 | 0.87 | 0.97 | rs28698859 | G | A | 0.04 | 110bp 3' of *ZNRD1-AS1* | intergenic |
| | | 0.91 | 1 | rs59854844 | G | A | 0.04 | *ZNRD1-AS1* | intergenic |
| 6 | 30001413 | 0.91 | 1 | rs72502509 | C | T | 0.04 | *ZNRD1-AS1* | intergenic |
| 6 | 30001656 | 0.91 | 1 | rs12665444 | G | C | 0.04 | *ZNRD1-AS1* | intergenic |
| 6 | 30002475 | 0.91 | 1 | rs10947043 | T | C | 0.04 | *ZNRD1-AS1* | intergenic |
| 6 | 30002561 | 0.91 | 1 | rs11755549 | A | G | 0.04 | *ZNRD1-AS1* | intergenic |
| 6 | 30002650 | 1 | 1 | rs11760176 | C | T | 0.04 | *ZNRD1-AS1* | intergenic |
| 6 | 30003151 | 0.91 | 1 | rs7742799 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30003257 | 0.91 | 1 | rs7762851 | T | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30003300 | 0.91 | 1 | rs7742712 | C | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30003514 | 0.87 | 0.97 | rs11964113 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30004196 | 0.91 | 1 | rs11754605 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30004615 | 0.91 | 1 | rs6933897 | T | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30004997 | 0.91 | 1 | rs6912001 | G | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30005193 | 0.91 | 1 | rs11965458 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30005410 | 0.91 | 1 | rs11965524 | T | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30006212 | 0.87 | 0.97 | rs73725950 | G | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30006306 | 0.87 | 0.97 | rs146541051 | CG | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30006569 | 0.91 | 1 | rs73414461 | G | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30007942 | 0.91 | 1 | rs73414465 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30007989 | 0.91 | 1 | rs3765608 | G | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30007998 | 0.91 | 1 | rs73725953 | A | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30008964 | 0.91 | 1 | rs12663252 | A | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30010250 | 0.91 | 1 | rs12661053 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30011866 | 0.91 | 1 | rs11753431 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30012381 | 0.84 | 0.93 | rs78453967 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30012413 | 0.84 | 0.93 | rs11753957 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30012463 | 0.91 | 1 | rs11753960 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30013222 | 0.84 | 0.93 | rs11754620 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30013591 | 0.91 | 1 | rs9261134 | G | C,T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30013956 | 0.91 | 1 | rs11759388 | A | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30014652 | 0.84 | 0.97 | rs11965585 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30014986 | 0.84 | 0.97 | rs11965659 | G | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30015638 | 0.84 | 0.97 | rs6902901 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30018365 | 0.84 | 0.97 | rs6931672 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30018724 | 0.84 | 0.97 | rs6932091 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30019414 | 0.84 | 0.97 | rs17187488 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30021187 | 0.84 | 0.97 | rs6928966 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30021411 | 0.84 | 0.97 | rs17193851 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30021912 | 0.84 | 0.97 | rs10947045 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30023157 | 0.84 | 0.97 | rs10807060 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30023760 | 0.84 | 0.97 | rs73725970 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30023796 | 0.84 | 0.97 | rs73416498 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30024837 | 0.84 | 0.97 | rs73416501 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30025737 | 0.84 | 0.97 | rs10550781 | ACACAC | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30026394 | 0.84 | 0.97 | rs35763150 | A | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30027247 | 0.84 | 0.97 | rs76290849 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30027961 | 0.84 | 0.97 | rs11967317 | G | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30030209 | 0.84 | 0.97 | rs7764237 | A | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30030247 | 0.84 | 0.97 | rs7746555 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30031232 | 0.84 | 0.97 | rs61230487 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30031451 | 0.84 | 0.97 | rs58555592 | A | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30032165 | 0.84 | 0.97 | rs2894006 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30032384 | 0.84 | 0.97 | rs112400819 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30032420 | 0.84 | 0.97 | rs113739823 | A | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30032946 | 0.84 | 0.97 | rs17187523 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30033196 | 0.84 | 0.97 | rs10947046 | T | G | 0.04 | *ZNRD1-AS1* | intronic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 30035041 | 0.84 | 0.97 | rs73420421 | T | C | 0.04 | *ZNRD1-AS1* | intergenic |
| 6 | 30036751 | 0.84 | 0.97 | rs58980848 | C | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30036837 | 0.84 | 0.97 | rs60265332 | A | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30036921 | 0.84 | 0.97 | rs58314476 | A | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30038011 | 0.84 | 0.97 | rs73420434 | A | T | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30038735 | 0.84 | 0.97 | rs73725984 | T | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30039339 | 0.84 | 0.97 | rs73420439 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30039934 | 0.84 | 0.97 | rs10947047 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30040207 | 0.84 | 0.97 | rs11962771 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30041484 | 0.84 | 0.97 | rs12661609 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30041887 | 0.84 | 0.97 | rs12664794 | A | G | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30043113 | 0.84 | 0.97 | rs60477028 | T | C | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30044563 | 0.84 | 0.97 | rs6940552 | G | A | 0.04 | *ZNRD1-AS1* | intronic |
| 6 | 30048520 | 0.84 | 0.97 | rs6457138 | A | C | 0.04 | *ZNRD1-AS1* | intronic |
| **Locus 9. Query SNP: rs1490387 and variants with $r^2$>= 0.8** | | | | | | | | | |
| 6 | 126337897 | 0.86 | 0.93 | rs6919397 | T | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126339209 | 0.86 | 0.93 | rs79950493 | T | TAGG | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126340008 | 0.86 | 0.93 | rs9388486 | T | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126340712 | 0.85 | 0.93 | rs9375435 | C | T | 0.46 | *CENPW* | intronic |
| 6 | 126343187 | 0.87 | 0.94 | rs17754780 | C | T | 0.45 | *CENPW* | intronic |
| 6 | 126345270 | 0.86 | 0.94 | rs6907898 | C | T | 0.45 | *CENPW* | intronic |
| 6 | 126345677 | 0.86 | 0.94 | rs200987420 | A | AAAT | 0.45 | *CENPW* | intronic |
| 6 | 126345678 | 0.86 | 0.94 | rs201988238 | A | AATAAC | 0.45 | *CENPW* | intronic |
| 6 | 126348333 | 0.86 | 0.94 | rs9401876 | C | A | 0.45 | *CENPW* | intronic |
| 6 | 126353208 | 0.87 | 0.94 | rs2045258 | A | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126357122 | 0.82 | 0.93 | rs9388487 | G | T | 0.47 | *RP11-73O6.4* | intergenic |
| 6 | 126357960 | 0.87 | 0.94 | rs9491624 | A | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126359343 | 0.87 | 0.94 | rs2326387 | G | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126360922 | 0.86 | 0.94 | rs35634111 | AT | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126362448 | 0.82 | 0.93 | rs9398803 | A | G | 0.47 | *RP11-73O6.4* | intergenic |
| 6 | 126364681 | 0.85 | 0.93 | rs2027299 | G | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126365367 | 0.87 | 0.94 | rs4895807 | C | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126377573 | 0.89 | 0.95 | rs9388489 | A | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126377904 | 0.89 | 0.95 | rs67126256 | TG | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126380821 | 0.87 | 0.94 | rs1361262 | T | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126382244 | 0.86 | 0.94 | rs9398804 | T | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126383649 | 0.85 | 0.95 | rs9388490 | C | T | 0.44 | *RP11-73O6.4* | intergenic |
| 6 | 126386699 | 0.84 | 0.92 | rs11307281 | CT | C | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126391101 | 0.9 | 0.95 | rs1578060 | G | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126405149 | 0.9 | 0.95 | rs9398805 | C | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126406266 | 0.83 | 0.93 | rs199553791 | T | TA | 0.44 | *RP11-73O6.4* | intergenic |
| 6 | 126406784 | 0.85 | 0.96 | rs4897179 | G | A | 0.44 | *RP11-73O6.4* | intergenic |
| 6 | 126406804 | 0.88 | 0.95 | rs4897180 | A | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126409397 | 0.89 | 0.94 | rs576049 | T | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126411948 | 0.8 | 0.92 | rs59698523 | T | C | 0.44 | *RP11-73O6.4* | intergenic |
| 6 | 126411994 | 0.82 | 0.9 | rs112166936 | A | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126419677 | 0.87 | 0.93 | rs34279746 | TTA | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126422941 | 0.9 | 0.95 | rs9321065 | G | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126423847 | 0.87 | 0.95 | rs145608233 | AATT | A | 0.44 | *RP11-73O6.4* | intergenic |
| 6 | 126423930 | 0.86 | 0.94 | rs2039735 | T | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126427347 | 0.91 | 0.96 | rs1538172 | A | G | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126430775 | 0.92 | 0.96 | rs9401881 | C | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126431652 | 0.92 | 0.96 | rs4565329 | C | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126431728 | 0.92 | 0.96 | rs1538170 | C | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126431738 | 0.92 | 0.96 | rs1538171 | C | G | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126432418 | 0.92 | 0.96 | rs4897181 | C | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126433909 | 0.92 | 0.96 | rs9401882 | A | G | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126434496 | 0.92 | 0.96 | rs4897182 | T | G | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126437394 | 0.92 | 0.96 | rs9375439 | C | G | 0.46 | *RP11-73O6.4* | intergenic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 126437645 | 0.92 | 0.96 | rs9398808 | G | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126437887 | 0.92 | 0.96 | rs9385399 | G | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126438441 | 0.92 | 0.96 | rs1415671 | T | G | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126438494 | 0.86 | 0.96 | rs149232167 | T | TG,TT | 0.44 | *RP11-73O6.4* | intergenic |
| 6 | 126439848 | 0.91 | 0.96 | rs2184968 | T | C | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126440082 | 0.92 | 0.96 | rs2152876 | G | A | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126443044 | 0.92 | 0.96 | rs9385400 | T | G | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126446365 | 0.91 | 0.96 | rs1361107 | A | C | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126446454 | 0.92 | 0.96 | rs1361108 | C | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126449878 | 0.92 | 0.96 | rs1337735 | T | A | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126449997 | 0.92 | 0.96 | rs1361109 | C | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126450541 | 0.93 | 0.96 | rs1572569 | G | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126452341 | 0.93 | 0.96 | rs2184967 | C | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126452434 | 0.93 | 0.96 | rs4559102 | G | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126454948 | 0.92 | 0.96 | rs9388494 | G | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126457454 | 0.93 | 0.96 | rs9375441 | G | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126460288 | 0.93 | 0.96 | rs4895808 | C | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126461311 | 0.92 | 0.96 | rs9388495 | G | A | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126463916 | 0.98 | 0.99 | rs9388496 | A | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126466961 | 0.97 | 0.99 | rs1844594 | G | A | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126467998 | 0.98 | 0.99 | rs9398809 | C | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126468326 | 0.98 | 0.99 | rs9385401 | C | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126475965 | 0.98 | 0.99 | rs9401883 | A | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126479998 | 0.99 | 0.99 | rs1159619 | C | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126481991 | 0.98 | 0.99 | rs9375442 | A | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126484109 | 0.99 | 0.99 | rs2050644 | A | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126485530 | 0.99 | 0.99 | rs4422634 | T | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126487491 | 0.97 | 0.99 | rs1120786 | T | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126488724 | 0.98 | 0.99 | rs9372839 | G | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126490782 | 0.98 | 0.99 | rs2326451 | A | T | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126494458 | 0.99 | 0.99 | rs9398810 | C | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126498553 | 0.99 | 0.99 | rs4418209 | T | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126501489 | 0.99 | 1 | rs9372840 | A | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126504691 | 1 | 1 | rs9401885 | C | G | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126504739 | 0.96 | 1 | rs72422730 | 8-mer | A | 0.44 | *RP11-73O6.4* | intergenic |
| 6 | 126505673 | 0.99 | 1 | rs2130603 | A | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126508835 | 0.99 | 1 | rs7738135 | G | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126509366 | 0.99 | 1 | rs4053271 | T | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126512937 | 1 | 1 | rs1490387 | C | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126513869 | 1 | 1 | rs9375446 | G | A | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126514509 | 1 | 1 | rs1490388 | C | T | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126516517 | 1 | 1 | rs1907067 | A | C | 0.45 | *RP11-73O6.4* | intergenic |
| 6 | 126517459 | 0.99 | 1 | rs1602278 | A | C,G | 0.46 | *RP11-73O6.4* | intergenic |
| 6 | 126530014 | 0.85 | 0.99 | rs1490384 | C | T | 0.49 | *RP11-73O6.4* | intergenic |
| **Locus 10. Query SNP: rs7022618 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 9 | 114813645 | 0.89 | 0.96 | rs10817677 | C | T | 0.14 | 7.5kb 5' of *TNFSF15* | intergenic |
| 9 | 114814199 | 0.89 | 0.96 | rs7865494 | C | T | 0.14 | 8.1kb 5' of *TNFSF15* | intergenic |
| 9 | 114814687 | 0.85 | 0.96 | rs7866221 | G | T | 0.15 | 8.6kb 5' of *TNFSF15* | intergenic |
| 9 | 114814957 | 0.85 | 0.96 | rs7866379 | C | T | 0.15 | 8.8kb 5' of *TNFSF15* | intergenic |
| 9 | 114816160 | 0.88 | 0.96 | rs10117997 | A | T | 0.15 | 10kb 5' of *TNFSF15* | intergenic |
| 9 | 114816811 | 0.8 | 0.96 | rs55897861 | A | 10-mer | 0.16 | 11kb 5' of *TNFSF15* | intergenic |
| 9 | 114818898 | 0.87 | 0.96 | rs12340243 | T | G | 0.15 | 13kb 5' of *TNFSF15* | intergenic |
| 9 | 114819380 | 0.86 | 0.96 | rs10982415 | G | A | 0.15 | 13kb 5' of *TNFSF15* | intergenic |
| 9 | 114819660 | 0.87 | 0.97 | rs2145929 | T | C | 0.15 | 14kb 5' of *TNFSF15* | intergenic |
| 9 | 114819807 | 0.87 | 0.97 | rs7873916 | A | G | 0.15 | 14kb 5' of *TNFSF15* | intergenic |
| 9 | 114823090 | 0.81 | 0.97 | rs7029554 | A | G | 0.16 | 17kb 5' of *TNFSF15* | intergenic |
| 9 | 114823185 | 0.82 | 0.97 | rs7032906 | C | T | 0.16 | 17kb 5' of *TNFSF15* | intergenic |
| 9 | 114824129 | 0.82 | 0.97 | rs1407306 | G | T | 0.16 | 18kb 5' of *TNFSF15* | intergenic |
| 9 | 114824742 | 0.86 | 0.97 | rs1419134 | A | G | 0.15 | 19kb 5' of *TNFSF15* | intergenic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 9 | 114825133 | 0.88 | 0.96 | rs12337233 | C | T | 0.15 | 19kb 5' of *TNFSF15* | intergenic |
| 9 | 114827294 | 0.81 | 0.96 | rs4262377 | G | T | 0.12 | 21kb 5' of *TNFSF15* | intergenic |
| 9 | 114827333 | 0.8 | 0.96 | rs10982417 | C | T | 0.12 | 21kb 5' of *TNFSF15* | intergenic |
| 9 | 114829725 | 0.8 | 0.96 | rs7468800 | C | A | 0.12 | 24kb 5' of *TNFSF15* | intergenic |
| 9 | 114831737 | 0.82 | 0.97 | rs7872350 | C | T | 0.12 | 26kb 5' of *TNFSF15* | intergenic |
| 9 | 114833900 | 0.83 | 0.98 | rs6478111 | C | T | 0.12 | 28kb 5' of *TNFSF15* | intergenic |
| 9 | 114838803 | 0.81 | 0.95 | rs10982420 | A | T | 0.13 | 33kb 5' of *TNFSF15* | intergenic |
| 9 | 114840016 | 0.8 | 0.94 | rs56235203 | A | G | 0.13 | 34kb 5' of *TNFSF15* | intergenic |
| 9 | 114840879 | 0.81 | 0.95 | rs10982421 | C | A | 0.13 | 35kb 5' of *TNFSF15* | intergenic |
| 9 | 114841503 | 0.81 | 0.95 | rs10982422 | T | C | 0.13 | 35kb 5' of *TNFSF15* | intergenic |
| 9 | 114841966 | 0.8 | 0.95 | rs1075074 | A | G | 0.13 | 36kb 5' of *TNFSF15* | intergenic |
| 9 | 114842234 | 0.8 | 0.95 | rs11560576 | A | T | 0.13 | 36kb 5' of *TNFSF15* | intergenic |
| 9 | 114842790 | 0.8 | 0.94 | rs11554257 | T | C | 0.13 | 37kb 5' of *TNFSF15* | intergenic |
| 9 | 114842969 | 0.81 | 0.95 | rs2418318 | T | C | 0.13 | 37kb 5' of *TNFSF15* | intergenic |
| 9 | 114843582 | 0.81 | 0.95 | rs10982423 | A | C | 0.13 | 37kb 5' of *TNFSF15* | intergenic |
| 9 | 114843864 | 0.82 | 0.96 | rs10982424 | G | A | 0.13 | 38kb 5' of *TNFSF15* | intergenic |
| 9 | 114843925 | 0.82 | 0.96 | rs10982425 | C | T | 0.13 | 38kb 5' of *TNFSF15* | intergenic |
| 9 | 114845338 | 0.82 | 0.97 | rs12237931 | G | A | 0.12 | 39kb 5' of *TNFSF15* | intergenic |
| 9 | 114845728 | 0.97 | 1 | rs1885385 | T | A | 0.14 | 40kb 5' of *TNFSF15* | intergenic |
| 9 | 114847443 | 0.82 | 0.97 | rs10491581 | T | A | 0.12 | 41kb 5' of *TNFSF15* | intergenic |
| 9 | 114847639 | 0.81 | 0.96 | rs16931895 | A | C | 0.12 | 42kb 5' of *TNFSF15* | intergenic |
| 9 | 114849633 | 0.83 | 0.99 | rs2093403 | T | C | 0.12 | 44kb 5' of *TNFSF15* | intergenic |
| 9 | 114850752 | 0.98 | 1 | rs10114224 | T | C | 0.14 | 43kb 3' of *TNFSF8* | intergenic |
| 9 | 114851416 | 0.98 | 1 | rs10982427 | T | C | 0.14 | 42kb 3' of *TNFSF8* | intergenic |
| 9 | 114852782 | 0.99 | 1 | rs4348576 | A | G | 0.14 | 41kb 3' of *TNFSF8* | intergenic |
| 9 | 114854276 | 1 | 1 | rs7022618 | T | C | 0.14 | 39kb 3' of *TNFSF8* | intergenic |
| 9 | 114855220 | 0.99 | 1 | rs10115186 | G | C | 0.14 | 39kb 3' of *TNFSF8* | intergenic |
| 9 | 114855286 | 0.83 | 0.99 | rs10982431 | G | A | 0.12 | 38kb 3' of *TNFSF8* | intergenic |
| 9 | 114855993 | 0.83 | 0.99 | rs79021241 | C | T | 0.12 | 38kb 3' of *TNFSF8* | intergenic |
| 9 | 114856377 | 0.99 | 1 | rs10119266 | T | A | 0.14 | 37kb 3' of *TNFSF8* | intergenic |
| 9 | 114856390 | 0.99 | 1 | rs10122171 | A | T | 0.14 | 37kb 3' of *TNFSF8* | intergenic |
| 9 | 114856435 | 0.99 | 1 | rs10116032 | C | T | 0.14 | 37kb 3' of *TNFSF8* | intergenic |
| 9 | 114856907 | 0.98 | 1 | rs16931913 | C | A | 0.14 | 37kb 3' of *TNFSF8* | intergenic |
| 9 | 114857329 | 0.99 | 1 | rs7043587 | T | C | 0.14 | 36kb 3' of *TNFSF8* | intergenic |
| 9 | 114857408 | 0.98 | 1 | rs7027268 | A | G | 0.14 | 36kb 3' of *TNFSF8* | intergenic |
| 9 | 114857479 | 0.99 | 1 | rs2418319 | T | C | 0.14 | 36kb 3' of *TNFSF8* | intergenic |
| 9 | 114857644 | 0.99 | 1 | rs2418320 | T | C | 0.14 | 36kb 3' of *TNFSF8* | intergenic |
| 9 | 114857767 | 0.83 | 0.99 | rs2418321 | C | T | 0.12 | 36kb 3' of *TNFSF8* | intergenic |
| 9 | 114858124 | 0.82 | 0.97 | rs10982433 | T | C | 0.12 | 36kb 3' of *TNFSF8* | intergenic |
| 9 | 114858435 | 0.82 | 0.97 | rs2145931 | T | C | 0.12 | 35kb 3' of *TNFSF8* | intergenic |
| 9 | 114858683 | 0.83 | 0.99 | rs149084342 | TC | T | 0.12 | 35kb 3' of *TNFSF8* | intergenic |
| 9 | 114859473 | 1 | 1 | rs10982434 | C | T | 0.14 | 34kb 3' of *TNFSF8* | intergenic |
| 9 | 114862450 | 0.83 | 0.99 | rs10982436 | G | C | 0.12 | 31kb 3' of *TNFSF8* | intergenic |
| 9 | 114863588 | 0.99 | 1 | rs7866387 | T | C | 0.14 | 30kb 3' of *TNFSF8* | intergenic |
| 9 | 114863619 | 0.83 | 0.99 | rs10982438 | G | A | 0.12 | 30kb 3' of *TNFSF8* | intergenic |
| 9 | 114864530 | 0.98 | 1 | rs7027093 | A | G | 0.14 | 29kb 3' of *TNFSF8* | intergenic |
| 9 | 114864911 | 0.85 | 0.99 | rs202132262 | CA | C | 0.12 | 29kb 3' of *TNFSF8* | intergenic |
| 9 | 114864914 | 0.97 | 0.99 | rs56136096 | T | A | 0.14 | 29kb 3' of *TNFSF8* | intergenic |
| **Locus 11. Query SNP: rs72758135 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 9 | 120862871 | 1 | 1 | rs72758135 | T | C | 0.13 | *PHF19* | intronic |
| 9 | 120864033 | 0.87 | 1 | rs10985064 | C | A | 0.12 | *PHF19* | intronic |
| 9 | 120864255 | 0.86 | 0.99 | rs12377227 | A | G | 0.12 | *PHF19* | intronic |
| | | 0.87 | 1 | rs148772466 | 8-mer | T | 0.12 | *PHF19* | intronic |
| 9 | 120864792 | 0.87 | 1 | rs72758138 | G | C | 0.12 | *PHF19* | intronic |
| 9 | 120865282 | 0.85 | 0.97 | rs62581711 | G | A | 0.12 | *PHF19* | intronic |
| 9 | 120866549 | 0.87 | 1 | rs10985066 | G | C | 0.12 | *PHF19* | intronic |
| 9 | 120867372 | 0.87 | 1 | rs10985067 | G | A | 0.12 | *PHF19* | intronic |
| 9 | 120867446 | 0.87 | 1 | rs10985068 | G | C | 0.12 | *PHF19* | intronic |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| **Locus 12. Query SNP: rs1360119 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 10 | 30401778 | 0.82 | 1 | rs75201195 | G | T | 0.02 | 27kb 5' of *MTPAP* | intergenic |
| 10 | 30415753 | 0.88 | 1 | rs116688297 | A | G | 0.02 | 18kb 5' of *MAP3K8* | intergenic |
| 10 | 30421514 | 0.82 | 0.94 | rs139462166 | G | T | 0.02 | 12kb 5' of *MAP3K8* | intergenic |
| 10 | 30422726 | 0.88 | 1 | rs143851974 | G | C | 0.02 | 11kb 5' of *MAP3K8* | intergenic |
| 10 | 30428353 | 0.82 | 0.94 | rs114738416 | G | A | 0.02 | 5.6kb 5' of *MAP3K8* | intergenic |
| 10 | 30434097 | 1 | 1 | rs8176952 | G | A | 0.02 | *MAP3K8* | 5'-UTR |
| 10 | 30434777 | 1 | 1 | rs1360119 | C | A | 0.02 | *MAP3K8* | intronic |
| 10 | 30435543 | 0.88 | 1 | rs76069125 | T | C | 0.02 | *MAP3K8* | intronic |
| 10 | 30436848 | 0.88 | 1 | rs79392497 | C | T | 0.02 | *MAP3K8* | intronic |
| 10 | 30440775 | 0.88 | 1 | rs141857700 | G | A | 0.02 | *MAP3K8* | intronic |
| **Locus 13. Query SNP: rs7092540 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 10 | 53092145 | 0.95 | 1 | rs951289 | C | T | 0.03 | 116kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53092368 | 1 | 1 | rs951287 | C | T | 0.03 | 117kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53092470 | 0.9 | 1 | rs951286 | C | T | 0.02 | 117kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53093470 | 0.95 | 1 | rs73350293 | T | A | 0.03 | 118kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53094196 | 1 | 1 | rs7092540 | G | A | 0.03 | 118kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53095751 | 0.9 | 1 | rs117932090 | G | A | 0.02 | 120kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53095970 | 0.9 | 1 | rs77554008 | T | C | 0.02 | 120kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53096806 | 0.9 | 1 | rs143715967 | 13-mer | T | 0.02 | 121kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53097291 | 0.9 | 1 | rs117797312 | C | T | 0.02 | 121kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53097965 | 0.9 | 1 | rs116906593 | C | T | 0.02 | 122kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53098299 | 0.9 | 1 | rs77174046 | A | G | 0.02 | 122kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53098634 | 0.9 | 1 | rs76456878 | C | T | 0.02 | 123kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53098965 | 0.9 | 1 | rs76515300 | T | C | 0.02 | 123kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53099452 | 0.9 | 1 | rs76657066 | G | A | 0.02 | 124kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53100607 | 0.9 | 1 | rs146597526 | G | A | 0.02 | 125kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53102644 | 0.9 | 1 | rs77627116 | G | A | 0.02 | 127kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53105027 | 0.9 | 1 | rs117659648 | C | T | 0.02 | 129kb 5' of *RP11-319F12.2* | intergenic |
| 10 | 53105878 | 0.9 | 1 | rs78263838 | T | C | 0.02 | 130kb 5' of *RP11-319F12.2* | intergenic |
| **Locus 14. Query SNP: rs59665078 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 12 | 128790563 | 1 | 1 | rs59665078 | T | C | 0.07 | 2.6kb 3' of *SLC15A4* | intergenic |
| **Locus 15. Query SNP: rs7149309 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 14 | 94467053 | 1 | 1 | rs7149309 | C | T | 0.01 | *SERPINA9* | intronic |
| **Locus 16. Query SNP: rs1802141 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 16 | 28956102 | 0.82 | 1 | rs73533880 | C | T | 0.01 | *NFATC2IP* | intronic |
| 16 | 28956367 | 0.82 | 1 | rs8047616 | C | T | 0.01 | *NFATC2IP* | intronic |
| 16 | 28957978 | 0.82 | 1 | rs111925530 | C | T | 0.01 | *RP11-264B17.2* | intronic |
| 16 | 28958196 | 0.82 | 1 | rs60603404 | T | A | 0.01 | *RP11-264B17.2* | intronic |
| 16 | 28961777 | 0.82 | 1 | rs113677814 | A | G | 0.01 | *RP11-264B17.2* | intronic |
| 16 | 28963337 | 0.82 | 1 | rs8044999 | A | G | 0.01 | *RP11-264B17.2* | intronic |
| 16 | 28970906 | 0.82 | 1 | rs8060015 | A | G | 0.01 | 3.3kb 5' of *SPNS1* | intergenic |
| 16 | 28971984 | 0.82 | 1 | rs148131850 | CT | C | 0.01 | 2.2kb 5' of *SPNS1* | intergenic |
| 16 | 28973283 | 0.82 | 1 | rs11863981 | G | C | 0.01 | 937bp 5' of *SPNS1* | intergenic |
| 16 | 28974410 | 0.82 | 1 | rs58714362 | T | A | 0.01 | *SPNS1* | intergenic |
| 16 | 28975676 | 0.82 | 1 | rs7184156 | G | A | 0.01 | *SPNS1* | intronic |
| 16 | 28976332 | 0.82 | 1 | rs7189583 | A | C | 0.01 | *SPNS1* | intronic |
| 16 | 28977491 | 0.82 | 1 | rs8044724 | G | A | 0.01 | *SPNS1* | intronic |
| 16 | 28979639 | 0.82 | 1 | rs9925291 | C | T | 0.01 | *SPNS1* | intronic |
| 16 | 28982048 | 0.82 | 1 | rs61747536 | C | T | 0.01 | *SPNS1* | synonymous |
| 16 | 28983736 | 0.82 | 1 | rs11859822 | T | C | 0.01 | *SPNS1* | intronic |
| 16 | 28985949 | 0.82 | 1 | rs41280846 | T | C | 0.01 | *RP11-264B17.3* | intronic |
| 16 | 28987287 | 0.82 | 1 | rs73535859 | C | T | 0.01 | *RP11-264B17.3* | intronic |
| 16 | 28987628 | 0.82 | 1 | rs7202093 | C | T | 0.01 | *RP11-264B17.3* | intronic |
| 16 | 28990688 | 1 | 1 | rs1802141 | A | G | 0.01 | *RP11-264B17.5* | 3'-UTR |

| Chromosome | Position (hg38) | LD ($r^2$) | LD (D') | Variant | Reference | Altered | EUR Frequency | GENCODE Genes | dbSNP Functional annotation |
|---|---|---|---|---|---|---|---|---|---|
| 16 | 28991788 | 1 | 1 | rs139964413 | AC | A | 0.01 | 1kb 3' of *LAT* | intergenic |
| 16 | 28993042 | 1 | 1 | rs112600506 | A | G | 0.01 | 2.3kb 3' of *LAT* | intergenic |
| 16 | 28993307 | 1 | 1 | rs111708809 | A | G | 0.01 | 2.5kb 3' of *LAT* | intergenic |
| 16 | 28994574 | 1 | 1 | rs7193414 | T | C | 0.01 | 3.8kb 3' of *LAT* | intergenic |
| **Locus 17. Query SNP: rs3848405 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 17 | 79119130 | 1 | 1 | rs3848405 | T | C | 0.01 | *RBFOX3* | intronic |
| 17 | 79119823 | 1 | 1 | rs139172998 | C | T | 0.01 | *RBFOX3* | intronic |
| 17 | 79124876 | 0.86 | 1 | rs73400195 | C | T | 0.01 | *RBFOX3* | intronic |
| **Locus 17. Query SNP: rs3848405 and variants with $r^2$ >= 0.8** | | | | | | | | | |
| 19 | 44895560 | 0.95 | 0.98 | rs2238681 | C | T | 0.42 | *TOMM40* | intronic |
| 19 | 44898409 | 0.95 | 0.98 | rs8106922 | A | G | 0.42 | *TOMM40* | Intronic |
| 19 | 44899220 | 0.96 | 0.98 | rs34878901 | C | T | 0.42 | *TOMM40* | Intronic |
| 19 | 44900155 | 0.85 | 0.99 | rs1160985 | C | T | 0.46 | *TOMM40* | Intronic |
| 19 | 44900601 | 0.85 | 0.99 | rs760136 | A | G | 0.46 | *TOMM40* | Intronic |
| 19 | 44901174 | 0.81 | 0.97 | rs741780 | T | C | 0.45 | *TOMM40* | Intronic |
| 19 | 44901715 | 0.84 | 0.98 | rs1038025 | T | C | 0.46 | *TOMM40* | Intronic |
| 19 | 44901805 | 0.85 | 0.99 | rs1038026 | A | G | 0.46 | *TOMM40* | Intronic |
| 19 | 44902264 | 0.95 | 0.98 | rs1305062 | G | C | 0.42 | *TOMM40* | Intronic |
| 19 | 44904531 | 0.85 | 0.99 | rs7259620 | G | A | 0.46 | 852bp 3' of *TOMM40* | Intergenic |
| 19 | 44907187 | 1 | 1 | rs769450 | G | A | 0.42 | *APOE* | Intronic |

**Table S6.** Differential expression of genes that are located within a 500 kilobase (kb) window around the candidaemia-associated SNPs upon *Candida* stimulation at 4 hours. Bolded genes show a log2 fold change > 1.5.

| Candidaemia SNP | Gene | P adjusted | Log2FoldChange |
|---|---|---|---|
| rs769450 | *BCL3* | 4.44E-36 | 1.11 |
| rs72758135 | *TRAF1* | 6.74E-58 | 1.73 |
| | *STOM* | 9.70E-10 | 1.25 |
| rs7149309 | *IFI27* | 1.04E-09 | 2.37 |
| | *SERPINA1* | 2.80E-25 | 1.12 |
| rs7022618 | *TNFSF15* | 7.72E-72 | 3.89 |
| | *TNFSF8* | 9.95E-21 | 1.01 |
| | *TNC* | 5.04E-08 | 2.28 |
| rs6748999 | *PROC* | 8.03E-03 | -1.22 |
| rs3848405 | *C1QTNF1* | 4.20E-09 | 2.59 |
| | *LGALS3BP* | 1.44E-19 | 1.49 |
| rs3766122 | *BLZF1* | 5.71E-17 | 1.33 |
| rs296537 | *LAD1* | 2.02E-11 | 2.32 |
| | *IGFN1* | 1.01E-06 | 1.19 |
| | *TNNT2* | 2.51E-04 | 1.65 |
| rs1802141 | *IL27* | 1.11E-11 | 2.21 |
| rs1360119 | *MAP3K8* | 1.67E-23 | 1.68 |
| rs12491812 | *CISH* | 7.92E-14 | 1.62 |

**Table S7.** Differential expression of genes that are located within a 500 kilobase (kb) window around the candidaemia-associated SNPs upon *Candida* stimulation at 24 hours. Bolded genes show a log2 fold change > 1.5.

| Candidaemia SNP | Gene | P adjusted | Log2FoldChange |
|---|---|---|---|
| rs6699706 | *AKR7A3* | 1.33E-02 | -1.36 |
| rs3766122 | *ATP1B1* | 1.50E-06 | 1.42 |
| rs296537 | *IGFN1* | 1.75E-20 | 5.14 |
| | *LAD1* | 6.45E-15 | 1.19 |
| rs6748999 | *PROC* | 1.16E-05 | -1.76 |
| | *MYO7B* | 9.41E-05 | 1.28 |
| rs12491812 | *SEMA3F* | 1.49E-07 | 2.70 |
| | *CISH* | 2.51E-26 | 2.45 |
| | *CACNA2D2* | 4.58E-16 | -1.14 |
| | *RASSF1* | 6.67E-33 | -1.09 |
| rs11760176 | *UBD* | 5.31E-03 | 2.18 |
| rs7022618 | *TNFSF15* | 9.00E-34 | 3.34 |
| rs72758135 | *STOM* | 1.25E-29 | 1.55 |
| rs1360119 | *ZNF438* | 3.90E-40 | 1.91 |
| | *MAP3K8* | 4.14E-45 | 1.39 |
| rs59665078 | *GLT1D1* | 7.24E-13 | 1.44 |
| rs7149309 | *IFI27* | 1.31E-08 | 2.94 |
| | *SERPINA1* | 4.87E-18 | 2.59 |
| | *SERPINA6* | 3.30E-02 | 1.72 |
| rs1802141 | *IL27* | 3.54E-10 | 3.60 |
| | *NUPR1* | 3.05E-03 | 2.19 |
| rs3848405 | *C1QTNF1* | 1.07E-35 | 6.37 |
| | *LGALS3BP* | 1.05E-06 | 1.50 |
| | *RBFOX3* | 4.78E-03 | -1.41 |
| rs769450 | *PVRL2* | 3.06E-17 | 2.51 |
| | *BCL3* | 3.32E-18 | 1.42 |
| | *APOE* | 1.15E-02 | -1.27 |
| | *RELB* | 9.48E-20 | 1.20 |

**Table S8.** All thirty-one prioritized susceptibility genes for candidaemia showed a strong enrichment for complement and coagulation pathways along with cytokine- and immune-related pathways based on KEGG and Reactome sources.

| p-value | q-value | Pathway | Source | Members/input/overlap |
|---|---|---|---|---|
| 0.0000001 | 0.0000064 | Complement and coagulation cascades - Homo sapiens (human) | KEGG | *SERPINA1; PROC; F5; C5; MBL2* |
| 0.0001401 | 0.0030831 | Staphylococcus aureus infection - Homo sapiens (human) | KEGG | *SELP; C5; MBL2* |
| 0.0002155 | 0.0031604 | Hemostasis | Reactome | *SELL; SERPINA1; F5; PROC; LAT; SELP* |
| 0.0004579 | 0.0039998 | Platelet degranulation | Reactome | *SERPINA1; SELP; F5* |
| 0.0005448 | 0.0039998 | Response to elevated platelet cytosolic Ca2+ | Reactome | *SERPINA1; SELP; F5* |
| 0.0007567 | 0.0039998 | Common Pathway of Fibrin Clot Formation | Reactome Reactome | *PROC; F5* |
| 0.0007760 | 0.0039998 | Platelet activation, signaling and aggregation | Reactome | *LAT; SERPINA1; SELP; F5* |
| 0.0007938 | 0.0039998 | Immune System | Reactome | *SELL; MAP3K8; TNFSF15; CISH; LAT; MBL2; C5; CD19* |
| 0.0008181 | 0.0039998 | Cell surface interactions at the vascular wall | Reactome | *SELL; PROC; SELP* |
| 0.0023813 | 0.0104777 | Formation of Fibrin Clot (Clotting Cascade) | Reactome | *PROC; F5* |
| 0.0047155 | 0.0188622 | Cytokine Signaling in Immune system | Reactome | *LAT; MAP3K8; CISH; TNFSF15* |
| 0.0085296 | 0.0312754 | Innate Immune System | Reactome | *LAT; MAP3K8; CD19; C5; MBL2* |

**Table S9.** Five additional candidaemia SNPs showed a moderate association with circulating cytokine levels as measured in serum from candidaemia patients. Cytokine levels were log transformed and statistical significance was tested with Kruskal Wallis test.

| SNP | P value (IFNγ) | P value (IL-6) | P value (IL-8) |
|---|---|---|---|
| rs296537 | 0.21 | 0.06 | 0.04 |
| rs7092540 | 0.01 | 0.20 | 0.24 |
| rs7149309 | 0.01 | 0.21 | 0.06 |
| rs1802141 | 0.00 | 0.07 | 0.04 |
| rs3848405 | 0.13 | 0.03 | 0.04 |

# CHAPTER 3

# A functional genomics approach to understanding human antifungal immune responses

Martin Jaeger, Vasiliki Matzaraki, Raul Aguirre-Gamboa, Mark S. Gresnigt, Xiaojing Chu, Melissa D. Johnson, Marije Oosting, Sanne P. Smeekens, Sebo Withoff, Iris Jonkers, John R. Perfect, Frank L. van de Veerdonk, Bart Jan Kullberg, Ramnik J. Xavier, Leo A.B. Joosten, Yang Li, Cisca Wijmenga, Mihai G. Netea, Vinod Kumar

# CHAPTER 3

## A functional genomics approach to understanding human antifungal immune responses

Martin Jaeger[1#], Vasiliki Matzaraki[2#], Raul Aguirre-Gamboa[2#], Mark S. Gresnigt[1], Xiaojing Chu[2], Melissa D. Johnson[3], Marije Oosting[1], Sanne P. Smeekens[1], Sebo Withoff[2], Iris Jonkers[2], John R. Perfect[3], Frank L. van de Veerdonk[1], Bart Jan Kullberg[1], Ramnik J. Xavier[4,5], Leo A.B. Joosten[1], Yang Li[2,8], Cisca Wijmenga[2,6,8], Mihai G. Netea[1,7,8], Vinod Kumar[1,2,8*]

[1] Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, 6525GA Nijmegen, the Netherlands.
[2] University of Groningen, University Medical Center Groningen, Department of Genetics, 9700RB Groningen, the Netherlands.
[3] Division of Infectious Diseases, Duke University Medical Center, Durham, North Carolina, USA and Department of Clinical Research, Campbell University School of Pharmacy, Buies Creek, North Carolina, USA.
[4] Broad Institute of Massachusetts Institute of Technology (MIT), Cambridge, MA, 02142, USA.
[5] Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, 02139, USA.
[6] K.G. Jebsen Coeliac Disease Research Centre, Department of Immunology, University of Oslo, Oslo, Norway.
[7] Human Genomics Laboratory, Craiova University of Medicine and Pharmacy, Craiova, Romania.
[#]These authors contributed equally to the manuscript
*Corresponding author: Vinod Kumar (v.kumar@umcg.nl)
[8]share last authorship

**Abstract**

Cytokines play a critical role in anti-fungal host defence, but a comprehensive approach to understanding the genetic architecture of cytokine responses against fungi and their role in fungal disease susceptibility is still lacking. To address this knowledge gap, we assessed the cytokine production capacity in different cellular systems following stimulation with three different opportunistic fungal pathogens: *Candida albicans* (yeast and hyphae), *Aspergillus fumigatus* and *Cryptococcus neoformans*. The inter-indi-

vidual variation pattern in cytokine production was specific for each fungal stimulus, and was similar between *C. albicans* yeast and hyphae. Using genome-wide genetic analysis we identified a novel significant locus that regulates IL-17A production upon stimulation with *C. albicans* hyphae. To explain susceptibility to systemic fungal infections, we performed the first candidaemia genome-wide association study (GWAS) and identified genes at 21 independent cytokine QTLs important for disease susceptibility. These loci were enriched for the lipid metabolism functional pathway. Our data suggest that genetic loci that affect cytokine responses to stimulation with fungi can partly explain an individual's fungal disease susceptibility.

Keywords: *C. albicans*, *A. fumigatus*, *C. neoformans*, cytokine-QTLs, susceptibility, candidaemia, genetic variants, lipid metabolism

## Introduction

Invasive fungal infections are an increasingly significant cause of sepsis in critically ill and immunocompromised patients and are associated with high morbidity and mortality. The majority of fungal-related deaths are caused by two fungal species acquired by the environment, namely, *Cryptococcus* and *Aspergillus* species, and by opportunistic *Candida* species, which are mainly part of the normal gut flora[1]. *Candida albicans* (*C. albicans*) is the most predominant fungal pathogen that causes both invasive and mucosal fungal infections. It is a dimorphic fungus as it grows both as yeast and filamentous cells, and causes approximately 250,000 new systemic infections on a yearly basis, leading to more than 50,000 deaths[2,3]. While *Candida* spp. co-exists as commensal in the human gastrointestinal tract and other mucous membranes, *Candida* can invade tissues and eventually blood circulation when epithelial barrier function is impaired and/or the host immune system is compromised, leading to systemic infections. From the bloodstream, *Candida* can disseminate to vital organs, eventually leading to organ failure followed by death.

Both innate and adaptive immune mechanisms are crucial for an effective host defence against fungi. The innate immune system represents the first line defence by recognition of the pathogen, followed by phagocytosis and/or killing. One of the defence mechanisms used by phagocytes is the use of reactive oxygen species (ROS) against various pathogens, including fungi[4]. Subsequently, innate immune cells contribute to the activation of adaptive immune response through antigen presentation and secretion of pro-inflammatory cytokines[5,6]. Cytokines in par-

ticular play a critical role during anti-*Candida* host immune defence, and recombinant cytokines have been proposed as adjunctive therapy in systemic *Candida* infection[7,8]. Administration of recombinant interferon (IFN)-γ also improved immune responses in patients with severe fungal infections[9,10].

In previous studies from the Human Functional Genomics Project (HFGP) consortium, we assessed the impact of genetic polymorphisms on cytokine production capacity in response to several pathogens, including fungi, in 500 healthy volunteers[11]. However, it remains unknown to what extent genetic variants that determine cytokine variability in response to *C. albicans*, can also affect susceptibility to systemic fungal infections, namely, candidaemia. In the present study, we first explored the genetic architecture of cytokine responses induced by the opportunistic fungal pathogen *C. albicans* and by the two other common fungal pathogens acquired by the environment *A. fumigatus* and *C. neoformans*. We also explored the differences in the cytokine responses induced by the two morphotypes of *C. albicans* (yeast and hyphae). In addition, we investigated the hypothesis whether susceptibility to candidaemia is explained by modulating the levels of pro-inflammatory cytokines. For this, we performed the first genome-wide association study (GWAS) of candidaemia on the largest candidaemia cohort published to date. Here we identified independent loci associated with candidaemia susceptibility to influence different cytokines produced from three different cell systems in response to *C. albicans* (Fig. 1). Finally, we tested whether these loci influence ROS production and *Candida* killing. Identifying genetic loci that influence susceptibility to systemic *Candida* infections by modulating cytokine levels will help us understand what drives susceptibility, a crucial step for the design of novel preventive and therapeutic strategies.

## Results

### Variations in cytokine immune response to opportunistic fungal pathogens in humans

To study the variations in cytokine responses to the three opportunistic fungal pathogens, we stimulated immune cells in three different cellular systems, whole blood (WB) for 48 hours, peripheral blood mononuclear cells (PBMCs) for 24 hours or 7 days and monocyte-derived macrophages for 24 hours with *C. albicans*, *A. fumifatus* and *C. neoformans* (Fig. 1)[11]. We used PBMCs, which is a mixed population of several cell types to capture interactions between different immune cell types. To study the interaction of different immune factors, such as between immune cell populations (granulocytes, monocytes and lymphocytes) and complement, antibodies and other serum factors, we also used WB stimulations. In addition to PBMCs and WB, we used monocyte-derived macrophages as these

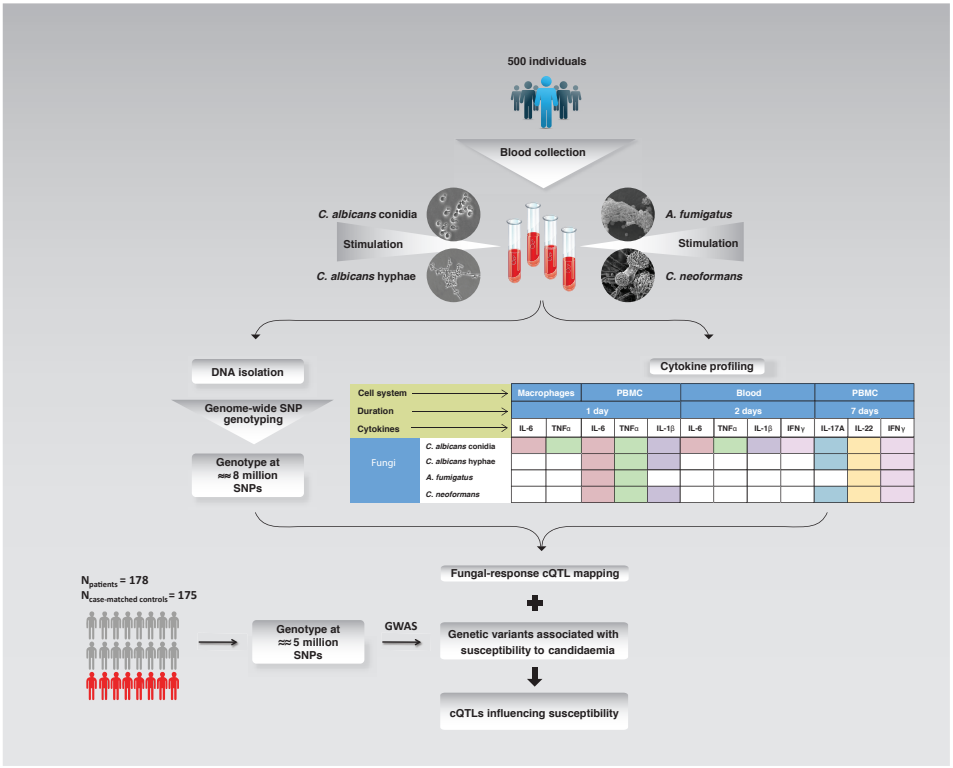



**Fig. 1.** Study overview. We collected blood samples from 500 healthy individuals in the 500FG cohort, isolated their DNA and genotyped it using the HumanOmniExpressExome-8 v1.0 SNP Chip. Imputation increased the genotype information to approximately 8 million SNPs. Whole blood or PBMCs or monocyte-derived macrophages were also used to perform a series of stimulation experiments with three different fungal species (*C. albicans*, *A. fumigatus* and *C. neoformans*) to profile the cytokines released in the three cell systems. Using the genotype information and cytokine measurements, we mapped fungal-response cQTLs. To identify whether cQTLs influence susceptibility to candidaemia, we performed a GWAS to test the association of genetic variants to disease susceptibility by using the largest candidaemia cohort available to date (178 cases and 175 case-matched controls).

cells are key effector cells in antifungal defence[5].

Unsupervised clustering of the levels of monocyte-derived cytokines (IL-1β, IL-6 and TNFα) and T-cell-derived cytokines (IFNγ, IL-17A and IL-22) upon stimulation with *C. albicans* yeast and hyphae in WB, PBMCs and monocyte-derived macrophages indicated that there is a clear distinction between WB, PBMCs and macrophage-derived cytokines (Fig. 2A). This observation suggests that additional variation in cytokine responses in WB may be explained by the contribution of different immune factors, such as different immune cell types (e.g neutrophils) or plasma factors, compared to the other two cell systems (PBMCs or monocyte-derived macrophages). In addition, clustering revealed a separation between mono-

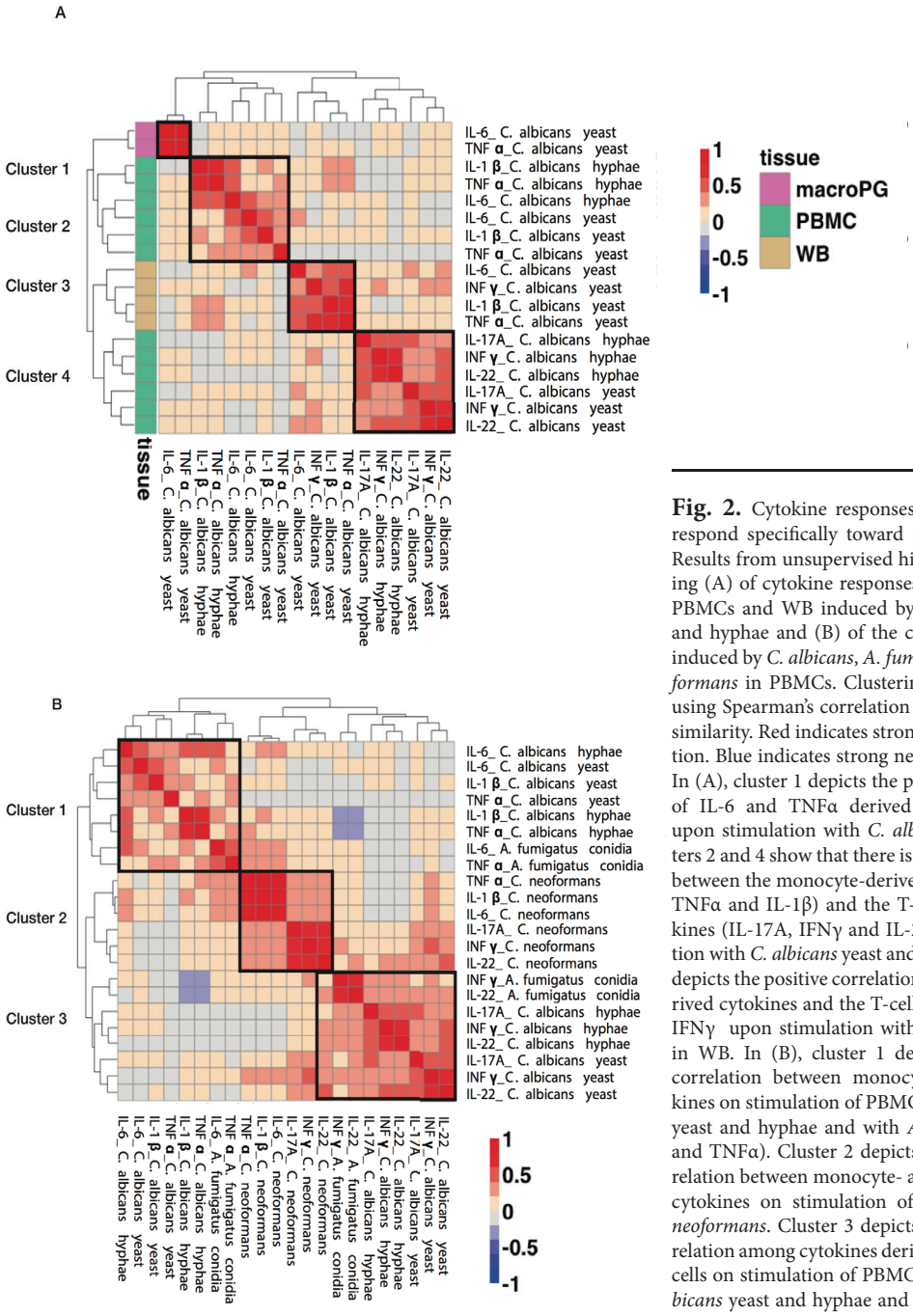



**Fig. 2.** Cytokine responses are organized to respond specifically toward fungal pathogens. Results from unsupervised hierarchical clustering (A) of cytokine responses in macrophages, PBMCs and WB induced by *C. albicans* yeast and hyphae and (B) of the cytokine responses induced by *C. albicans*, *A. fumigatus* and *C. neoformans* in PBMCs. Clustering was performed using Spearman's correlation as the measure of similarity. Red indicates strong positive correlation. Blue indicates strong negative correlation. In (A), cluster 1 depicts the positive correlation of IL-6 and TNFα derived by macrophages upon stimulation with *C. albicans* yeast. Clusters 2 and 4 show that there is a clear distinction between the monocyte-derived cytokines (IL-6, TNFα and IL-1β) and the T-cell-derived cytokines (IL-17A, IFNγ and IL-22) upon stimulation with *C. albicans* yeast and hyphae. Cluster 3 depicts the positive correlation of monocyte-derived cytokines and the T-cell-derived cytokine IFNγ upon stimulation with *C. albicans* yeast in WB. In (B), cluster 1 depicts the positive correlation between monocyte-induced cytokines on stimulation of PBMCs with *C. albicans* yeast and hyphae and with *A. fumigatus* (IL-6 and TNFα). Cluster 2 depicts the positive correlation between monocyte- and T–cell-derived cytokines on stimulation of PBMCs with *C. neoformans*. Cluster 3 depicts the positive correlation among cytokines derived from T-helper cells on stimulation of PBMCs with both *C. albicans* yeast and hyphae and with *A. fumigatus* (IL-22, IL-17A and IFNγ). macroPG: macrophages, PBMC: Peripheral blood mononuclear cells, WB: whole blood.

cyte-derived and T-cell-derived cytokine responses within the PBMC fraction. Furthermore, the production of cytokines induced by the two morphotypes of *C. albicans* (yeast and hyphae) was generally strongly correlated.

To assess the correlation between cytokine responses induced by *C. albicans*, *A. fumigatus* and *C. neoformans*, we performed unsupervised clustering of cytokines produced in PBMCs in response to all three fungi. Clustering was mainly dependent on specific fungal species rather than between different cytokines, and we observed distinct clusters for *C. albicans*, *A. fumigatus* and *C. neoformans* (Fig. 2B). At the level of host anti-fungal response, this observation further substantiates the concept that immune responses are organized to respond specifically to defined fungal pathogens rather than to respond uniformly to any infection with the production of the same cytokines[11]. Interestingly, we observed a stronger positive correlation between cytokines induced by *A. fumigatus* and *C. albicans* compared to those induced by *C. neoformans* (Fig. 2B). In contrast, only the T-cell-derived cytokines induced by *C. neoformans* were more similar to the T-cell-derived cytokine responses induced by *C. albicans* yeast (Fig. 2B). Of note, the distinct cluster between monocyte-derived and T-cell-derived cytokines was weaker for *C. neoformans*-induced cytokine responses compared to the *C. albicans*-induced responses (Fig. 2B).

## Genetic variation has a role in heterogeneity of anti-*Candida c*ytokine responses

We observed that cytokine levels show a significant variation upon stimulation with *C. albicans* yeast and hyphae in all our stimulation cell systems compared to stimulation with the medium control (Figure S1). We have previously shown that non-genetic host factors such as age and gender affect cytokine responses to fungal stimulation[12], while cell counts had only a minor effect[13]. Additionally, we identified three genome-wide significant SNPs that affect *C. albicans*-yeast-induced cytokines (cytokine-QTLs or cQTLs) and two genome-wide significant SNPs that affect *C. neoformans*-induced cytokines, indicating that genetics have an impact on cytokine responses upon fungal stimulation[13].

Given that clustering analysis revealed a strong positive correlation between cytokine responses induced by *C. albicans* yeast and hyphae (Fig. 2), we hypothesized that the cytokine responses to each of these morphotypes are regulated by similar genetic factors. To test this hypothesis, we investigated whether the same genetic variants determine the cytokine responses induced by the two morphotypes of *C. albicans*. For this, we first mapped *C. albicans*-hyphae-response cQTLs by correcting for age, gender and cell counts using genome-wide genotypes from the 500FG cohort. A novel genome-wide significant cQTL ($P < 5x10^{-8}$) at chromosome 16p12.2 (Fig. 3A) was identified. This top cQTL, rs13333993 was associ-

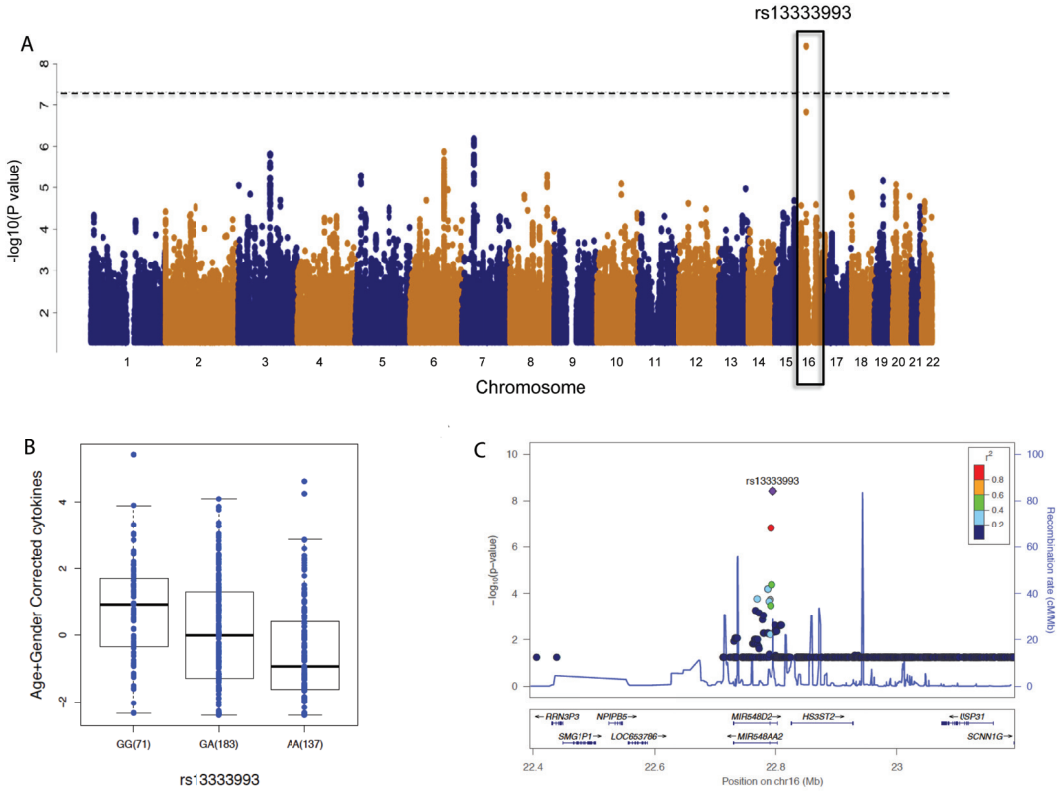



**Fig. 3.** Genome-wide QTL mapping identifies a novel *C. albicans* hyphae response cQTL. (A) Manhattan plot showing the genome-wide QTL mapping results for *C. albicans* hyphae-induced IL-17A levels. Y-axis represents the –log10P values of cQTLs. Their chromosomal positions are shown on the X-axis. Horizontal dashed line represents the genome-wide significance threshold for association ($P < 5x10^{-8}$). (B) Box plot showing the association of the genotypes of SNP rs13333993 with IL-17A levels upon stimulation with *C. albicans* hyphae in PBMCs for 7 days. The number of individuals per genotype is shown in parentheses below the box plot. (C) Regional association plot showing the –log10P for SNPs centered on the most strongly associated signal (rs13333993, purple diamond) associated with IL-17A levels upon *C. albicans* hyphae stimulation in PBMCs for 7 days.

ated with IL-17A levels induced by *C. albicans* hyphae (Fig. 3B), and this SNP is located in close proximity to the *HS3ST2* gene that codes for the heparan sulfate glucosamine 3-O-sulfotransferase 2 protein (Fig. 3C). Further investigation revealed that this SNP does not influence the gene expression of *HS3ST2*, as shown in publicly available eQTL datasets[14,15].

Given that unsupervised clustering of the cytokine responses induced by *C. neoformans* and *C. albicans* yeast indicated a positive correlation within the T-cell-derived cytokines induced by these two fungi (Fig. 2B), we investigated whether this correlation is driven by shared genetic components. For this, we test-

ed if independent *C. albicans*-response cQTLs ($P < 9.99x10^{-4}$) could be associated with cytokine levels induced by *C. neoformans* with a nominal significance ($P < 0.05$). We found a significant enrichment of shared cQTLs from T-cell-derived cytokines induced by *C. albicans* yeast and by *C. neoformans* compared to monocyte-derived cytokines (Figure S2). Of note, this enrichment remained significant across the different P-value thresholds used to define cQTLs (Figure S2).

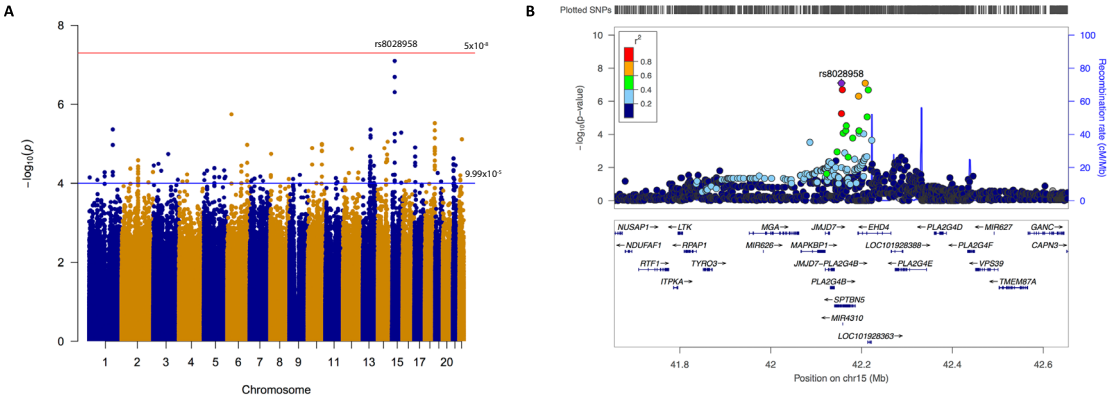



**Fig. 4.** *PLA2G4B* locus is associated with candidaemia susceptibility. (A) Manhattan plot of SNPs associated with candidaemia susceptibility identified by GWAS. GWAS revealed a locus strongly associated with susceptibility to candidaemia at chromosome 15 ($P = 7.97x10^{-8}$, OR = 0.4002). Y-axis represents –log10P values of SNPs. X-axis shows chromosomal positions. Blue line is the suggestive threshold for association ($P < 9.99x10^{-5}$). (B) Regional association plot for the candidaemia-associated SNP, rs8028958 (purple diamond), which mapped to the *JMJD7-PLAS2G4B* locus at chromosome 15. Each dot represents a SNP, and the linkage disequilibrium of neighboring SNPs with the top SNP is color-coded. Y-axis represents –log10P values of SNPs. X-axis shows chromosomal positions.

## Candidaemia GWAS identifies the *PLA2G4B* locus association with susceptibility

To identify genetic variants associated with susceptibility to candidaemia, we genotyped 178 well described candidaemia patients[65] and 175 case-matched controls of European ancestry on a genome-wide scale and imputed the genotype data using the Human Reference Consortium (HRC). After quality control filtering per SNP and sample, we obtained 161 patients and 152 case-matched controls for whom we had genotype information on 5,326,313 SNPs variants, which were tested for disease association. Finally, this GWAS identified 235 SNPs in 70 independent loci that show suggestive disease associations ($P < 9.99x10^{-5}$). Among these loci, a SNP on chromosome 15, rs8028958, showed the strongest association with susceptibility to candidaemia ($P = 7.97x10^{-8}$, OR = 0.4002) (Fig. 4A). This top

associated SNP is located within the intronic region of the *SPTBN5* gene (Fig. 4B). Previous eQTL analysis showed that SNP rs11070352, a proxy SNP to rs8028958 ($r^2$ = 0.9 and D' = 0.99), affects the expression of *EHD4* ($P$ = 7.76x10$^{-27}$) and *JMJD7-PLA2G4B* ($P$ = 2.63x10$^{-18}$) in WB[15]. In addition, eQTL data from GTEx project also indicated the effect of this SNP on different genes in different tissues including *PLA2G4B*, suggesting its strong regulatory function[16]. *JMJD7-PLA2G4B* encodes a fusion protein known as JMJD7-PLA2G4B and is a paralog gene of *PLA2G4B*[17]. In addition, we observed that differential expression analysis upon *C. albicans* stimulation in PBMCs prioritized *PLA2G4B* as a plausible candidaemia susceptibility gene (Figure S3A). *PLA2G4B* gene expression was significantly decreased ($P$ 1.02x10$^{-13}$) upon stimulation of PBMCs with *C. albicans* after 24 hours compared to RPMI stimulated PBMCs.

Given the strong association of rs8028958 with susceptibility to candidaemia ($P$ 7.97x10$^{-8}$) and the significant differential expression of *PLA2G4B* in response to *C. albicans* in PBMCs ($P$ 1.02x10$^{-13}$), we performed *in vitro* functional experiments to validate the role of *PLA2G4B* in anti-*Candida* host immune defence. To investigate the importance of the *PLA2G4B* locus in the host immune response against *C. albicans*, human primary monocytes were isolated and specific *PLA2G4B* siRNA or control siRNA were used to silence the gene prior to stimulation with *C. albicans*. In the following experiment, cytokine production capacity of TNFα, IL-1β, IL-6 and IL-1Ra, known to play an important role in anti-*Candida* host immune defence[5], were measured. Comparing cytokine levels of non-targeting control siRNA with *PLA2G4B* siRNA revealed a decrease in IL-6 and TNFα production in the targeted group (Figure S3B). Finally, targeted silencing of *PLA2G4B* also led to a reduction of the capacity of monocytes to kill *C. albicans* yeast (Fig. 5), suggesting a role for *PLA2G4B* in host defence against *C. albicans*.

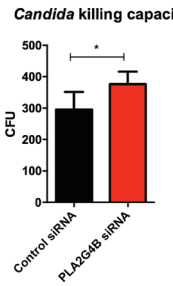



**Fig. 5**. *C. albicans* killing capacity of primary monocytes upon targeted silencing of the *PLA2G4B* gene. A significant reduction of the killing capacity of monocytes was observed upon siRNA silencing of the *PLA2G4B* gene compared to control siRNA. Data from three experiments with at least three donors per experiment was pooled. CFU: colony forming units.

## SNPs modulating cytokine levels associate with candidaemia susceptibility

To investigate whether SNPs modulating cytokine levels upon *C. albicans* stimulation are associated with candidaemia susceptibility – which would demonstrate the importance of inherited cytokine production capacity for susceptibility to infection – we tested whether genetic variants associated with susceptibility to candidaemia influence cytokine levels in response to *C. albicans* yeast or hyphae

using the results of the previous cQTL mapping. We found 75 SNPs from 21 independent candidaemia-susceptibility loci (Table 1) that modulate cytokine levels upon *C. albicans* stimulation in any of the cell systems (PBMCs, WB, and macrophages). Twenty SNPs modulate multiple cytokines upon stimulation with both *Candida* morphotypes (Table S1). These observations suggest that susceptibility can be partly explained by defects in cytokine pathways, and that there are also other alternative mechanisms involved in candidaemia susceptibility.

Moreover, from these loci, we extracted genes in a 500 Kb window upstream and downstream of each cQTL and prioritized genes that could determine susceptibility by modulating cytokine levels (Table 1). For gene prioritization, we used publicly available eQTL mapping datasets from blood[14,15]. Our prioritization identified a number of genes, including *ALOX15B*, *GGT5*, *HSD17B4*, J*MJD7-PLA2G4B*, *AGPAT1*, *TTC7B* and *ALOXE3*, known to be involved in different lipid metabo-

**Table 1.** SNPs modulating cytokines in response to *C. albicans* yeast and hyphae ($P < 0.05$) in one of the cell systems (blood, PBMCs and monocyte-derived macrophages) show an association with susceptibility to candidaemia ($P < 9.99\text{x}10^{-5}$).

| Locus | SNPs | Chr | Base Pair | *C. albicans* stimulation | Cytokine | Cell system | Time | P-value | | OR | Gene(s)[f] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Susceptibility | cQTL | | |
| 1 | rs1483198 | 1 | 4717391 | yeast | IL-6 | Blood | 48 hr | $7.10\text{x}10^{-5}$ | $1.62\text{x}10^{-2}$ | 2.322 | *AJAP1*[a], *LOC284661*[b] |
| 2 | rs35523523 | 2 | 241581438 | hyphae | TNFα | PBMC | 24 hr | $5.17\text{x}10^{-5}$ | $3.91\text{x}10^{-2}$ | 2.202 | *AQP12B*[b], *AQP12A*[b], *DUSP28*[b] |
| | | | | yeast | TNFα | Blood | 48 hr | $5.17\text{x}10^{-5}$ | $3.27\text{x}10^{-2}$ | | |
| 3 | rs6758021 | 2 | 127305291 | yeast | IFNγ | PBMC | 7 days | $7.45\text{x}10^{-5}$ | $4.53\text{x}10^{-2}$ | 0.3728 | *GYPC*[b] |
| 4 | rs28378632 | 4 | 62743831 | yeast | IL-6 | PBMC | 24 hr | $8.80\text{x}10^{-5}$ | $4.44\text{x}10^{-2}$ | 0.499 | *LPHN3*[b], *LPHN3-AS1*[b] |
| 5 | rs6872016 | 5 | 25359184 | yeast | IL1-β | Blood | 48 hr | $4.76\text{x}10^{-5}$ | $1.53\text{x}10^{-2}$ | 0.3853 | |
| 6 | rs4895365 | 5 | 118641096 | yeast | IFNγ | PBMC | 7 days | $6.94\text{x}10^{-5}$ | $2.32\text{x}10^{-3}$ | 0.4643 | *HSD17B4*[c], *TNFAIP8*[d] |
| | | | | hyphae | IL-1β | | 24 hr | | $2.22\text{x}10^{-2}$ | | |
| | | | | yeast | IL-22 | | 7 days | | $2.89\text{x}10^{-3}$ | | |
| 7 | rs492899 | 6 | 31933518 | yeast | IL-17A | PBMC | 7 days | $1.79\text{x}10^{-6}$ | $1.08\text{x}10^{-2}$ | 0.1839 | *AGPAT1*[c], *HSPA1B*[c] |
| | | | | | TNFα | | 24 hr | | $2.06\text{x}10^{-2}$ | | |
| 8 | rs11155859 | 6 | 152874252 | yeast | IFNγ | PBMC | 7 days | $1.06\text{x}10^{-5}$ | $4.93\text{x}10^{-2}$ | 0.4829 | *MYCT1*[b], *SYNE1*[b], *VIP*[b] |
| 9 | rs2709952 | 7 | 82879055 | yeast | IL-6 | PBMC | 24 hr | $6.71\text{x}10^{-5}$ | $2.09\text{x}10^{-2}$ | 2.028 | *PCLO*[b], *SEMA3E*[b], *MIR7976*[b] |
| 10 | rs10107521 | 8 | 142546574 | yeast | IFNγ | Blood | 48 hr | $7.83\text{x}10^{-5}$ | $7.96\text{x}10^{-3}$ | 3.55 | *MROH5*[b], *PTP4A3*[b], *GPR20*[b] |
| 11 | rs2725008 | 8 | 4478647 | yeast | IL-6 | Macrophages | 24 hr | $6.17\text{x}10^{-5}$ | $1.59\text{x}10^{-2}$ | 0.4301 | *CSMD1*[b] |
| | | | | | TNFα | | | | $1.02\text{x}10^{-2}$ | | |
| 12 | rs1945666 | 11 | 81325329 | hyphae | TNFα | PBMC | 24 hr | $7.37\text{x}10^{-5}$ | $2.58\text{x}10^{-2}$ | 1.988 | *MIR4300*[b], *LOC101928989*[b] |
| 13 | rs6563046 | 13 | 79057998 | hyphae | IL-1β | PBMC | 24 hr | $4.35\text{x}10^{-6}$ | $4.94\text{x}10^{-2}$ | 2.12 | *POU4F1*[b], *RNF219*[b] |
| | | | | yeast | TNFα | | | | $1.91\text{x}10^{-3}$ | | |
| 14 | rs7322708 | 13 | 75381199 | yeast | IL-1β | Blood | 48 hr | $2.08\text{x}10^{-5}$ | $3.59\text{x}10^{-2}$ | 2.337 | *LINC00347*[b], *CTAGE11P*[b] |
| | | | | hyphae | IL-22 | PBMC | 7 days | | $3.43\text{x}10^{-2}$ | | |
| 15 | rs7337751 | 13 | 71895852 | yeast | TNFα | Macrophages | 24 hr | $7.61\text{x}10^{-5}$ | $6.57\text{x}10^{-3}$ | 2.046 | |
| 16 | rs71423384 | 14 | 70442987 | hyphae | IL-1β | PBMC | 24 hr | $1.29\text{x}10^{-5}$ | $1.88\text{x}10^{-2}$ | 0.3945 | *SLC8A3*[a], *SMOC1*[b], *SRSF5*[b] |
| 17 | rs1294517 | 14 | 90910208 | yeast | IL-1β | PBMC | 24 hr | $2.43\text{x}10^{-5}$ | $3.46\text{x}10^{-2}$ | 2.355 | *TTC7B*[a], *KCNK13*[a] |
| | | | | | IL-6 | | | | $2.89\text{x}10^{-2}$ | | |
| 18 | rs1113333 | 15 | 42194217 | yeast | IFNγ | Blood | 48 hr | $6.13\text{x}10^{-5}$ | $3.92\text{x}10^{-2}$ | 0.5196 | *EHD4*[c,d], *JMJD7-PLA2G4B*[c] |
| 19 | rs3027232 | 17 | 8022065 | yeast | IL-17A | PBMC | 7 days | $9.27\text{x}10^{-5}$ | $4.41\text{x}10^{-2}$ | 0.4685 | *C17orf44*[c], *TMEM107*[c], *TMEM88*[d], *ALOX15B*[a], *ALOXE3*[a] |
| 20 | rs72987764 | 18 | 76395325 | yeast | IL-17A | PBMC | 7 days | $3.01\text{x}10^{-5}$ | $4.20\text{x}10^{-2}$ | 0.3844 | *SALL3*[b], *ATP9B*[b] |
| | | | | | TNFα | | 24 hr | | $2.79\text{x}10^{-3}$ | | |
| 21 | rs113413 | 22 | 24292264 | hyphae | IL-17A | PBMC | 7 days | $8.04\text{x}10^{-5}$ | $3.76\text{x}10^{-2}$ | 2.034 | *AP000350.4*[c], *AP000350.5*[c], *AP000350.6*[c], *AP000351.10*[c], *DDT*[c], *DDTL*[c], *GSTT1*[c], *KB-226F1.1*[c], *KB-226F1.2*[c], *GGT5*[a] |

[a] Gene differentially expressed in response to *C. albicans* stimulation in PBMCs at 24 hours
[b] Gene in close proximity to the cytokine QTL is shown
[c] Expression QTL mapping in blood shows a correlation between cytokine QTL SNP and the expression of that gene
[d] Gene differentially expressed in response to *C. albicans* stimulation in PBMCs at 4 hours
[e]Twenty-one independent *C. albicans* yeast or hyphae-response cQTLs associated with candidaemia susceptibility ($P < 0.05$) in any of the cell systems: WB, PBMCs and monocyte-derived macrophages.
[f]Genes in a 500 Kb window upstream and downstream of each cQTL were prioritized from each locus based on gene expression data of PBMCs in response to *C. albicans* for 4 and 24 hours and eQTL data from whole blood.

lism processes, a finding which was also confirmed by pathway enrichment analysis (Table S2).

Lastly, as ROS production is an important aspect of phagocyte-mediated host responses, we tested whether any of the 21 susceptibility loci that influence cytokine levels has an impact on *Candida* killing through influencing ROS production. For this, we mapped QTLs using genotypes and measurements of ROS production and *Candida* killing of 137 samples from an independent population-based cohort (200FG)[18,19]. We found two loci that have an impact on *Candida* killing by influencing ROS production (Fig. 6A and B). In particular, cQTL (rs2725008), an intronic variant of *CSMD1* gene at chromosome 8 (Fig. 6A), influences *Candida*

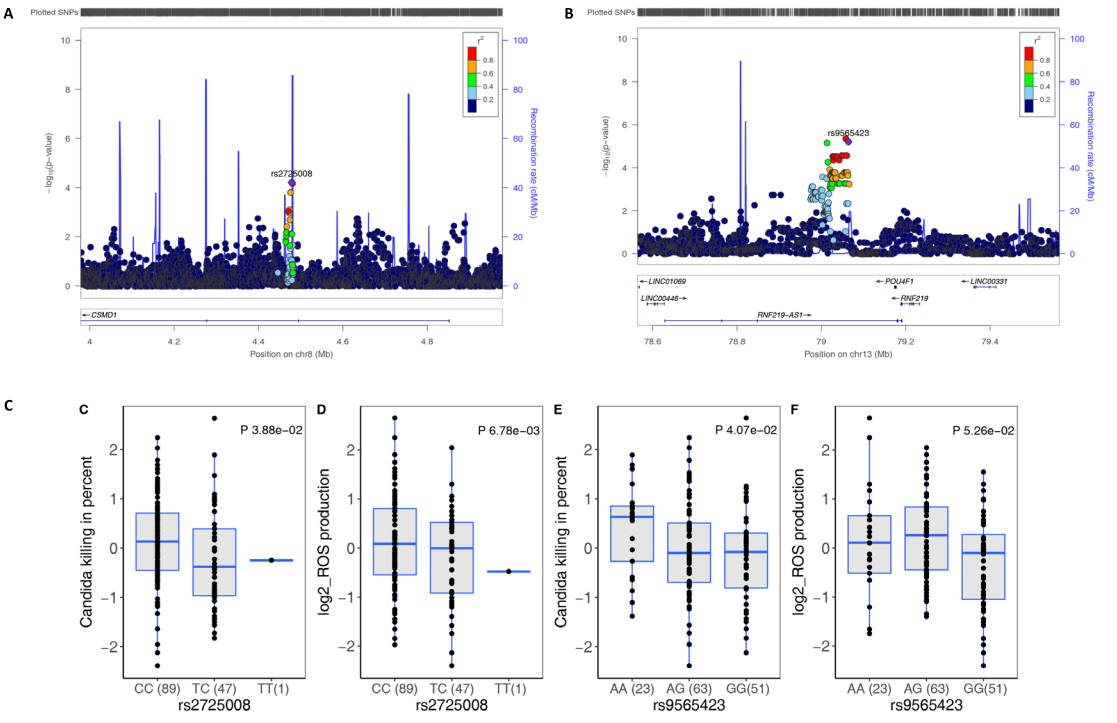



**Fig. 6.** Related to Table 1. Two susceptibility loci influencing cytokine levels have an impact on *Candida* killing by influencing ROS production. Regional association plots for (A) rs2725008 and (B) rs9565423 showing the –log10P of association for susceptibility of SNPs centered on the most strongly associated signal (purple diamond). Each dot represents a SNP, and different colors of dots indicate varying linkage disequilibrium with the top signal. Box plots showing the association of genotypes between SNP rs2725008 at chromosome 8 with (C) *Candida* killing (in percentage) and (D) ROS production, and between SNP rs9565423 at chromosome 13 with (E) *Candida* killing (in percentage) and (F) ROS production. The data shown were generated using 137 samples from 200FG cohort for whom both genotype and measurements of ROS production and *Candida* killing were available. For normalization, log2 transformation was used for the measurements of ROS production and rank-based inversed normal transformation for the percentages of *Candida* killing.

killing ($P$ 3.88x10$^{-2}$) through affecting ROS production (RLU/sec) ($P$ 6.78x10$^{-3}$) (Fig. 6 C to D). In addition, a second variant rs9565423 at chromosome 13 (Fig. 6B), a proxy ($r^2$ >= 0.8) to susceptibility variant rs6563046 influencing cytokines, found to influence *Candida* killing ($P$ 4.07x10$^{-2}$) and ROS production ($P$ 5.26x10$^{-2}$) with a P nominal significance (Fig. 6 E to F).

These findings indicate that these two loci have a functional role in *Candida* killing by influencing ROS production and that cytokines may influence *Candida* killing through other mechanisms, independent of ROS production.

## Discussion

In a given at-risk population not all patients develop infections, which indicates that there is a strong genetic influence on individual susceptibility to different infections[20]. Although systemic fungal infections result in high morbidity and mortality in developed countries, genetic association studies that identify risk genes for systemic fungal infections are extremely challenging because of the difficulty of enrolling large numbers of case-matched controls. To circumvent these challenges, we identified genetic variants that affect human cytokine responses to common opportunistic fungal pathogens (*C. albicans*, *A. funigatus* and *C. neoformans*) by studying the variability of cytokine production in PBMCs isolated from healthy volunteers from 500FG cohort. Further, we explored the genetic variation that determines cytokine responses induced specifically by the two morphotypes of *C. albicans* (yeast and hyphae) in PBMCs. To explore whether genetic variation influencing cytokine levels in response to *C. albicans* explains susceptibility to candidaemia, we intersected the *Candida*-response cQTLs with genetic variants identified by the first GWAS exploring susceptibility to candidaemia using the largest cohort available to date.

Our study reveals that the inter-individual variations in cytokine production were similar for *C. albicans* yeast and hyphae. Through genetic analysis we identified a significant novel genome-wide genetic locus that regulates IL-17A production upon stimulation with *C. albicans* hyphae. We also found a significantly higher contribution of shared genetics in the regulation of T-cell-derived cytokine levels in response to both *C. albicans* and *C. neoformans*, a relationship that was not present in monocyte-derived cytokines. Finally, by integrating *Candida*-response cQTLs with SNPs associated with candidaemia-susceptibility in our GWAS, we identified 21 independent cQTLs important in determining susceptibility to candidaemia.

To cause infection, fungi need to cross skin and/or mucosal barriers, penetrate

and invade underlying tissues and at the same time resist human immune defences[21]. To achieve this, these fungi adapt to changing environments by developing different escape strategies. The human host responds to these strategies by organizing the host cytokine response in a pathogen-specific manner, shown by the distinct clusters of correlation we observed for the cytokine production induced by *A. fumigatus*, *C. albicans* or *C. neoformans* (Fig. 2). It has been shown that *C. albicans* hyphae (associated with tissue invasion) has a different cell wall composition compared to yeast cells (associated with colonization), indicating differential host recognition of hyphae and yeast as well as explaining differences in the virulence of these two morphological forms[22,23]. Interestingly, the high correlation between hyphae- and yeast-induced cytokine responses suggests that cytokine responses to these two morphotypes are regulated by shared genetic and non-genetic factors.

An important finding of the present study is the identification of novel genetic factors that influence *C. albicans*-induced cytokine production while also impacting susceptibility to candidaemia. Moreover, we found a cQTL with genome-wide significance for *C. albicans* hyphae-induced IL-17A production. The gene located closest to this variant is *HS3ST2*, which codes for the heparan sulfate glucosamine 3-O-sulfotransferase 2 protein. *HS3ST2* is an important gene involved in macrophage polarization, and we have previously identified this gene as one of the differentially regulated genes by the TLR locus genotype[24]. These results indicate that some of the cQTLs influencing *Candida*-induced cytokines are downstream of TLR-signaling, which is in line with earlier literature demonstrating the role of TLRs for pattern recognition of *C. albicans*[5]. What remains to be investigated is the precise mechanism through which *HS3ST2* modulates IL-17A production.

One of the most important aspects of our study was our demonstration in an unbiased manner of the strong impact of cQTLs on susceptibility to candidaemia. As shown in Table 1, there are 21 independent cQTLs that modulate susceptibility to candidaemia. Among these polymorphisms, 20 SNPs modulate cytokines in response to both *C. albicans* morphotypes (Table S1). Two loci out of the 21 found to have an impact on *Candida* killing by influencing ROS production, which is one of the major defence mechanism against different pathogens, including fungi [18,19,25]. Of particular interest, rs2725008 is an intronic variant in *CSMD1* gene, which encodes a novel complement-regulatory protein highly expressed in epithelial tissues and in the central nervous system[26], suggesting that the complement system may have a role in determining susceptibility to *Candida* infections. Notably, the complement system is a major component of the innate immune system and plays a critical role in defending against invading pathogens[27]. Moreover, genes from susceptibility loci that influence cytokines found to be enriched for those involved in lipid and lipoprotein metabolism processes (Table S2). Many metabolic changes including lipid synthesis are shown to be important in regu-

lating cytokine secretion to coordinate immune response[28–30]. In mouse studies, hyperlipoproteinemia increased the susceptibility to systemic candidiasis due to an increased fungal outgrowth in their organs[31], while a beneficial effect was seen in rats infected with gram-negative bacteria[32]. These data were in agreement with human studies where infusion of lipoproteins enhances the growth of *C. albicans* in the plasma of volunteers[33]. Although the lipid profiles differ between humans and mice, both studies suggest that hyperlipidemia have deleterious effects by enhancing the growth of *C. albicans* in both species. Therefore, it needs to be tested whether the candidaemia-associated SNPs that we have identified in this study influence cytokine levels through affecting lipids and lipoprotein levels. Nevertheless, our data demonstrate the power of a systems-based approach for identifying the most important host defence mechanisms against an infection.

Finally, we have identified a novel genome-wide genetic association of candidaemia with polymorphisms in *PLA2G4B*, a gene encoding the cytosolic phospholipase A2 protein that selectively hydrolyses phospholipids to release lysophospholipids and fatty acids. It has been suggested that PLA2G4B-F enzymes may control phospholipid and/or arachidonate metabolism in a tissue specific manner[34]. However, the role of *PLA2G4B* has never been studied in the context of infectious diseases. Interestingly, specific down-regulation of *PLA2G4B* using siRNA resulted in impairment of the candidacidal properties of monocyte-derived macrophages, confirming its importance in the host response against *C. albicans*.

In conclusion, our study provides an important resource for the understanding of genetic factors in determining cytokine production induced by opportunistic fungal pathogens and susceptibility to candidaemia. The novel strategy we have taken demonstrates the strength of a systems genomics approach to identifying novel risk genes and pathways for fungal infections. In the future, systematic integration of such datasets may also be useful for other infectious diseases and could ultimately lead to the development of novel therapies and treatment strategies that target these genes and pathways.

## Materials and methods

### Study populations

*Population-based cohorts*. To understand the variation in cytokine production in humans in response to *C. albicans in vitro*, we used two independent population-based cohorts, 500FG and 200FG, composed of healthy individuals of Western European ancestry from the Human Functional Genomics Project (HFGP, see www.humanfunctionalgenomics.org). The 500FG cohort is composed of 534 well-characterized healthy individuals, 237 males and 296 females all between 18

and 75 years old. The 200FG cohort is composed of ~200 healthy individuals from the 'Geldersch Landschap', 'Hoge Veluwe', 'Twickel', and 'Kroondomein het Loo' in the Netherlands, is 77% male and 23% female[13] and has an age range from 23 to 73 years old.

*Candidaemia cohort*. To identify genetic variants associated with candidaemia susceptibility, we performed the first GWAS analysis of a well-described candidaemia cohort with corresponding disease-matched controls[35]. In total, 178 candidaemia cases and 175 disease-matched controls of European ancestry were tested for disease association. The demographic and clinical characteristics of the candidaemia cohort have been described previously[35]. The patients were enrolled between January 2003 and January 2009. Patients must have had at least one positive blood culture for a *Candida* species, with the majority of them infected by *C. albicans*[35]. Case-matched controls were recruited from the same hospital wards as candidaemia cases so that co-morbidities and clinical risk factors for infection were similar. Both candidaemia cases and case-matched controls were enrolled after informed written consent was given at the Duke University Hospital (DUMC, Durham, North Carolina, USA). The study was approved by the institutional review boards at the Duke University Hospital.

**PBMC collection and stimulation experiments**

Venous blood from the cubital vein of healthy volunteers was drawn in EDTA tubes after obtaining written informed consent. PBMC isolation was performed as previously described [36]. In short, the PBMC fraction was obtained using density centrifugation in Ficoll-Paque (Pharmacia Biotech). Cells were then washed twice in PBS and re-suspended in RPMI medium supplemented with gentamicin 10 mg/mL, L-glutamine 10 mM and pyruvate 10 mM. PBMCs were counted and re-suspended in a concentration of $5x10^6$ cells/mL. $5x10^5$ PBMCs were added in 100 µL to round-bottom 96-well plates (Greiner) and incubated with 100 µL of stimulus (RPMI, heat killed *Candida albicans* yeast $1x10^6$/mL or *Candida albicans* hyphae $1x10^5$/mL). After 24 hours or 7 days in the presence of 10% human serum, the supernatants were collected and stored at −20°C. Cytokines were measured using different ELISA assays. *C. albicans* yeast cells of the strain ATCC MYA-3573, UC820 were used for the stimulation experiments. To generate hyphae, live yeast forms of *Candida* were grown for 24 hours at 37°C in RPMI (Gibco-BRL), adjusted to pH 6.4 by using hydrochloric acid. After 24 hours, more than 95% of the yeast cells were grown into hyphae, which were checked by microscope. Hyphae were heat killed for 45 min at 98°C and resuspended in RPMI to a hyphal inoculum size that originated from $10^6$/ml yeast cells (referred to as $10^6$/mL hyphae).

### Macrophage differentiation and stimulation

After PBMC isolation, purification of monocytes was performed by hyper-osmotic Percoll density gradient centrifugation. $1.5x10^5$ cells were seeded in a 96-well flat bottom plate (Greiner). After 24 hours of attachment, 25 nM of either non-targeting siRNA control pool (D-001810-10-05) or *PLA2G4B* (L-187552-01-0005) targeting siRNA (Smartpool, Thermo Scientific) were added for 24 hours. After the transfection period, cells were stimulated for 24 hours with $1x10^6$/mL *Candida* yeast. After the stimulation period, supernatants were collected and stored at −20°C. Cytokines were measured using different ELISA assays.

### Whole-blood stimulation experiments

100 µL of fresh heparin blood was added to a 48-well plate and then stimulated with 400 µL stimulus (either RPMI or *C. albicans* $1x10^6$/mL) for 48 hr at 37°C and 5% $CO_2$. Supernatants were collected and stored at –20°C until used for ELISA.

### Cytokine measurements

Frozen supernatants from PBMCs, monocytes and WB stimulation experiments were measured for IL-1β, IL-6, TNFα, IL-22, IL-17A, and IFNγ using commercial available ELISA kits (PeliKine Compact or R&D Systems). All ELISAs were carried out according to manufacturer protocol. Data were analyzed using GraphPad prism v5.0. Data are shown as means ± SEM. $P < 0.05$ = *, $P < 0.01$ = ** and $P < 0.001$ = ***.

### ROS production

Induction of reactive oxygen species (ROS) was measured by oxidation of luminol (5-amino-2,3, dihydro-1,4-phtalazinedione). $5x10^5$ PBMCs of healthy volunteers were resuspended in HBSS and put in dark 96 wells plates. Cells were exposed to $1x10^6$/mL *C. albicans* yeast (UC820, heat killed) together with 20µl of 1mM luminol (final concentration 50 µM). Chemiluminescence was measured in BioTek Synergy HTreader at 37°C for 1 hour with intervals of 2.23 minutes. The relative luminescence units per second (RLU/sec) within the area under the curve (AUC) were plotted against time and analysed by using Graphpad Prism v5.0. $P < 0.05$ = *, $P < 0.01$ = ** and $P < 0.001$ = ***.

### Monocyte fungal killing assays

For the monocyte siRNA experiments, $1.5x10^5$ cells were directly seeded into 96-well flat bottom plates. After 24 hours of siRNA treatment, $1x10^6$/mL live *C. albicans* yeast was added for 24h of incubation at 37°C. Following incubation, well contents were serial diluted in sterile water and plated on Sabouraud agar for counting of colony forming units. Data was stratified regarding genotypes for

SNP rs8028958 in *PLA2G4B*. Data from three experiments with at least three donors per experiment was pooled and data was analyzed using Graphpad Prism v5.0. $P < 0.05$ = *, $P < 0.01$ = ** and $P < 0.001$ = ***.

**Variance analysis and cytokine clustering**

We first log-transformed raw cytokine levels and filtered out cytokine measurements showing little/no variation across individuals as previously described[24]. We used the Levene's test to check the equality of variance of cytokine levels before and after stimulation. We performed unsupervised hierarchical clustering using Spearman's correlation as the measure of similarity. All statistical analyses regarding the cytokine data were performed in R (https://www.R-project.org/).

**Genotyping, quality control and imputation of the 500FG cohort**

DNA obtained from the 500FG cohort was genotyped using the commercially available Illumina HumanOmniExpressExome-8 v1.0 SNP chip. Genotype calling was performed using Opticall 0.7.0 using default settings[37]. We applied quality control per sample to exclude samples with a call rate ≤ 0.99 and quality control per SNP to exclude variants with a Hardy-Weinberg equilibrium ≤ 0.0001, a call rate ≤ 0.99 and a minor allele frequency (MAF) ≤ 0.001. We identified 17 ethnic outliers by merging multi-dimensional scaling plots of samples with 1000 Genomes data, and these outliers were excluded from further analysis[24]. The quality control filters resulted in a dataset of 483 samples containing genotype information on 518,980 variants for further imputation. The strands and variants-identifiers were aligned to the reference Genome of the Netherlands (GoNL, Genome of the Netherlands Consortium, 2014[38]) dataset using Genotype Harmonizer[39]. The data were phased using SHAPEIT2 v2 using GoNL as a reference panel[40,41]. We selected SNPs that showed an INFO score ≥ 0.8 upon imputation for further cytokine QTL mapping.

**Genotyping, quality control and imputation of the candidaemia cohort**

Isolated DNA obtained from case and control samples was genotyped using the commercially available SNP chips, HumanCoreExome-12 v1.0 and HumanCoreExome-24 v1.0 BeadChip from Illumina (https://www.illumina.com). Genotype calling was performed using Optical 0.7.0 using the default settings[37]. Strands of variants were aligned and identified against the 1000 Genome reference panel using Genotype Harmonizer[39]. We then imputed the samples on the Michigan imputation server using the human reference consortium as a reference panel, and we filtered variants with an $R^2$ of 0.3 for imputation quality[42]. After excluding imputed variants with a MAF < 0.10 and variants within the HLA region, we applied quality control per sample and removed 15 individuals (5 cases and 10 controls) due to excess/reduced heterozygosity and cryptic relatedness. Next,

we applied SNP quality control filters to exclude SNPs with (a) a MAF of < 0.05 and (b) a Hardy-Weinberg equilibrium of $P < 1x10^{-6}$ in control samples only. We also identified 25 ethnic outliers (12 cases and 13 controls) by multidimensional scaling analysis (performed in PLINK on the N x N matrix of genome-wide IBS pairwise distance) of candidaemia patients and disease-matched controls for the two first principle components (Figure S4A). The quantile-quantile plot that shows the distribution of the observed –log10 for the whole-genome SNPs against the theoretical distribution of expected –log10 indicated no or little evidence of population stratification (Figure S4B). The genomic inflation factor based on median chisq was equal to 1 ($\lambda = 1$). Quality control filters resulted in a dataset of 161 cases and 152 disease-matched control samples containing genotype information on 5,326,313 SNP variants for further testing for association with candidaemia susceptibility.

### Cytokine QTL mapping

We selected 442 individuals from 500FG for whom both genotype and cytokine data were available. We obtained cell count data measured by FACS for total lymphocytes, T cells, B cells, monocytes and NK-cells from 487 individuals from the 500FG cohort. For cytokine QTL mapping, we used 409 samples with genotype, cytokine and cell-count data. Of these, 17 samples were excluded due to genetic differences. To correct the cytokine distributions for QTL mapping, we used a linear model adjusting for age, gender and cell-count. Raw cytokine levels were first log-transformed and the correlation between cytokine production and genotype was tested by linear regression model with age and gender as co-variables using the R-package Matrix-eQTL. Since the stimulation-cytokine combinations cannot be regarded as completely independent, the total number of independent tests among all phenotypes will be much less than the number of phenotypes measured, and therefore correcting for the number of SNPs tested for each trait should be sufficient. We considered $P < 5x10^{-8}$ to be the threshold for significant cQTLs. *C. albicans* response QTLs ($P < 0.05$) were intersected with SNPs associated with susceptibility to candidaemia showing a P-value between $9.99x10^{-5}$ and $5x10^{-8}$ to identify cQTLs associated with susceptibility.

### QTL mapping of ROS production and *Candida* killing

We selected 137 healthy individuals from 200FG for whom both genotype and measurements of ROS production and *Candida* killing (in percentages) were available. We applied a log2 transformation and rank-based inverse normal transformation on raw measurements of ROS production and *Candida* killing respectively and the correlation between ROS production and *Candida* killing with genotypes was tested by linear regression model with age and gender as co-variables using the R-package Matrix-eQTL. We considered $P < 0.05$ to be the threshold for significant QTLs.

### Pathway enrichment analysis

We performed gene over-representation analysis using ConsensusPathDB-human database (CPDB; http://cpdb.molgen.mpg.de/). The analysis was done using the default settings, where pre-defined lists of functionally associated genes (pathways, Gene Ontology categories and neighborhood-based entity sets) are tested for over-representation in a user-specified list based on the hypergeometric test. For each of the predefined sets, a P-value was calculated according to the hypergeometric test based on the number of physical entities present in both the pre-defined set and the user-specified list of physical entities. P-values were corrected for multiple testing using the false discovery rate method and are available as q-values. To identify independent cQTL loci, linkage disequilibrium (LD) SNP pruning of cQTLs using genotypes extracted from the 500FG cohort was performed based on pairwise genotypic correlation using PLINK 1.90 (https://www.cog-genomics.org/plink2), with one of a pair of SNPs removed if LD was > 0.2. Next, we extracted genes in a window of ±500 kb from the twenty-one independent loci and filtered the genes using differential gene expression data of PBMCs in response to *C. albicans* for 4 and 24 hours. Once this had all been done, we finally ran the over-representation analysis.

### Genome-wide case-control association analysis

The associations between the GWAS-identified variants and candidaemia susceptibility were tested by Fisher's exact test using PLINK 1.9 (www.cog-genomics.org/plink/1.9/)[43]. The genomic inflation factor ($\lambda = 1$) indicated that there was no population stratification effect (Figure S5). A P value significance threshold of $< 5.0x10^{-8}$ was set to call genome-wide significant associations. We considered SNPs with a P-value between $9.99x10^{-5}$ and $5x10^{-8}$ as variants showing a suggestive association with candidaemia susceptibility.

### List of abbreviations

*C. albicans*: *Candida albicans*; *A. fumigatus*: *Aspergillus fumigatus*; *C. neoformans*: *Cryptococcus neoformans*; SNP: single nucleotide polymorphism, cQTL: cytokine-quantitative trait locus

### Declarations

### Ethics approval and consent to participate

All volunteers gave written informed consent prior to blood donations and the study was approved by the ethical review board of Radboud University Nijmegen. The Human Functional Genomics Project (HFGP) was approved by the Ethical Committee of Radboud University Nijmegen, the Netherlands (no. 42561.091.12).

Experiments were conducted according to the principles expressed in the Declaration of Helsinki and venous blood samples were drawn after written informed consent was obtained.

### Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request

### Competing interests

The authors declare no competing interests.

### Funding

This work was supported by a European Research Council (ERC) Consolidator Grant [FP/2007-2013/ERC grant 2012-310372] and a Netherlands Organization for Scientific Research (NWO) Spinoza prize grant [NWO SPI 94-212] to M.G.N., an ERC Advanced grant [FP/2007-2013/ERC grant 2012-322698] and an NWO Spinoza prize grant [NWO SPI 92-266] to C.W., a European Union Seventh Framework Program grant (EU FP7) TANDEM project [HEALTH-F3-2012-305279] to C.W. and V.K., and a Research Grant [2017] of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) to V.K. Y.L. and M.O. were supported by a VENI grant (863.13.011 and 016.176.006) from the Netherlands Organization for Scientific Research (NWO). V.M. is supported by a PhD scholarship from Graduate School of Medical Sciences, University of Groningen, the Netherlands.

### Author's contributions

Conceptualization:M.J.,V.M.,R.A.G.,C.W.,M.G.N.,V.K.;Methodology: L.A.B.,Y.L.,C.W.,M.G.N.,V.K.;Analysis:V.M.,M.J.,R.A.G.,M.S.G.,M.O.,S.P.S .,S.W.I.J.,Data collection: M.J.,M.D.J.,M.G.,M.O.,S.P.S.,J.R.P.,F.L.V.,B.J.K.,R. J.X.,L.A.B., Data analysis: M.J., V.M., R.A.G., X.C., V.K., Y.L.;Writing: M. J.,V.M,M.G.N.,V.K.;Project administration: M.G.N.,C.W., V.K.; Funding acquisition:M.G.N.V.K.,M.O.,Y.L. All authors read and approved the manuscript.

### Acknowledgements

The authors thank all volunteers from the 500 and 200 Functional Genomics cohort of the Human Functional Genomics Project for participation in the study. The authors would like to thank K. McIntyre for editing the final text.

# References

1. Brown, G. D. et al. Hidden killers: human fungal infections. Sci. Transl. Med. 4, 1–9 (2012).
2. Boyce, K. J. & Andrianopoulos, A. Fungal dimorphism: The switch from hyphae to yeast is a specialized morphogenetic adaptation allowing colonization of a host. FEMS Microbiol. Rev. 39, 797–811 (2015).
3. Campion, E. W., Kullberg, B. J. & Arendrup, M. C. Invasive Candidiasis. N. Engl. J. Med. 373, 1445–1456 (2015).
4. Paiva, C. N. & Bozza, M. T. Are reactive oxygen species always detrimental to pathogens? Antioxid. Redox Signal. 20, 1000–1037 (2014).
5. Netea, M. G., Joosten, L. A., van der Meer, J. W., Kullberg, B. J. & van de Veerdonk, F. L. Immune defence against Candida fungal infections. Nat Rev Immunol 15, 630–642 (2015).
6. Netea, M. G., Brown, G. D., Kullberg, B. J. & Gow, N. a R. An integrated model of the recognition of Candida albicans by the innate immune system. Nat. Rev. Microbiol. 6, 67–78 (2008).
7. van de Veerdonk, F. L., Kullberg, B. J. & Netea, M. G. Adjunctive immunotherapy with recombinant cytokines for the treatment of disseminated candidiasis. Clin. Microbiol. Infect. 18, 112–119 (2012).
8. Armstrong-James, D. et al. Immunotherapeutic approaches to treatment of fungal diseases. The Lancet Infectious Diseases e393–e402 (2017). doi:10.1016/S1473-3099(17)30442-5
9. Delsing, C. E. et al. Interferon-gamma as adjunctive immunotherapy for invasive fungal infections: a case series. BMC Infect. Dis. 14, 166 (2014).
10. Buddingh, E. P. et al. Interferon-gamma immunotherapy in a patient with refractory disseminated candidiasis. Pediatr Infect Dis J 34, 1391–1394 (2015).
11. Li, Y. et al. Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. Nat. Med. 22, 952–60 (2016).
12. ter Horst, R. et al. Host and environmental factors influencing individual human cytokine responses. Cell 167, 1111–1124.e13 (2016).
13. Li, Y. et al. Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. Nat. Med. 22, 952–960 (2016).
14. GTEx Consortium, Gte. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–60 (2015).
15. Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–43 (2013).
16. Carithers, L. J. et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreserv. Biobank. 13, 311–319 (2015).
17. Cheng, Y., Wang, Y., Li, J., Chang, I. & Wang, C.-Y. A novel read-through transcript JMJD7-PLA2G4B regulates head and neck squamous cell carcinoma cell proliferation and survival. Oncotarget 8, 1972–1982 (2017).
18. Aratani, Y. et al. Critical role of myeloperoxidase and nicotinamide adenine dinucleotide phosphate-oxidase in high-burden systemic infection of mice with Candida albicans. J. Infect. Dis. 185, 1833–1837 (2002).
19. Frohner, I. E., Bourgeois, C., Yatsyk, K., Majer, O. & Kuchler, K. Candida albicans cell surface superoxide dismutases degrade host-derived reactive oxygen species to escape innate immune surveillance. Mol. Microbiol. 71, 240–252 (2009).
20. Ottenhoff, T. H. M., Verreck, F. A. W., Hoeve, M. A. & Van De Vosse, E. Control of human host immunity to mycobacteria. Tuberculosis 85, 53–64 (2005).
21. Köhler, J. R., Casadevall, A. & Perfect, J. The spectrum of fungi that infects humans. Cold Spring Harb. Perspect. Med. 5, a019273 (2015).
22. Cheng S-C, van de Veerdonk FL, Lenardon M, et al. The dectin-1/inflammasome pathway is responsible for the induction of protective T-helper 17 responses that discriminate between yeasts and hyphae of Candida albicans. J. Leukoc. Biol. 90, 357–366 (2011).
23. Heinsbroek, S. E. M., Brown, G. D. & Gordon, S. Dectin-1 escape by fungal dimorphism. Trends Immunol. 26, 352–354 (2005).
24. Li, Y. et al. A functional genomics approach to understand variation in cytokine production in humans. Cell 167, 1099–1110.e14 (2016).

25. Gazendam, R. P. et al. Two independent killing mechanisms of Candida albicans by human neutrophils: Evidence from innate immunity defects. Blood 124, 590–597 (2014).
26. Kraus, D. M. et al. CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. J. Immunol. 176, 4419–4430 (2006).
27. Zipfel, P. F. Complement and immune defense: From innate immunity to human diseases. Immunol. Lett. 126, 1–7 (2009).
28. Dimeloe, S., Burgener, A. V., Grählert, J. & Hess, C. T-cell metabolism governing activation, proliferation and differentiation; a modular view. Immunology 150, 35–44 (2017).
29. Lachmandas, E. et al. Microbial stimulation of different Toll-like receptor signalling pathways induces diverse metabolic programmes in human monocytes. Nat. Microbiol. 2, 16246 (2016).
30. Netea MG, D. C. More than Inflammation: Interleukin-1β Polymorphisms and the Lipid Metabolism. J. Clin. Endocrinol. Metab. 96, 1279–1281 (2011).
31. Netea, M. G. et al. Hyperlipoproteinemia enhances susceptibility to acute disseminated Candida albicans infection in low-density-lipoprotein-receptor-deficient mice. Infect. Immun. 65, 2663–2667 (1997).
32. Read, T. E. et al. Triglyceride-rich lipoproteins prevent septic death in rats. J. Exp. Med. 182, 267–272 (1995).
33. Netea, M. G. et al. Infusion of lipoproteins into volunteers enhances the growth of Candida albicans. Clin. Infect. Dis. 28, 1148–51 (1999).
34. Long, J. Z. & Cravatt, B. F. The metabolic serine hydrolases and their functions in mammalian physiology and disease. Chem. Rev. 111, 6022–6063 (2011).
35. Kumar, V. et al. Immunochip SNP array identifies novel genetic variants conferring susceptibility to candidaemia. Nat. Commun. 5, 4675 (2014).
36. Oosting, M. et al. Borrelia-induced cytokine production is mediated by spleen tyrosine kinase (Syk) but is Dectin-1 and Dectin-2 independent. Cytokine 76, 465–472 (2015).
37. Shah, T. S. et al. OptiCall: A robust genotype-calling algorithm for rare, low-frequency and common variants. Bioinformatics 28, 1598–1603 (2012).
38. Collection, S. & Genome, T. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. 46, 1–95 (2014).
39. Deelen, P. et al. Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. BMC Res. Notes 7, 901 (2014).
40. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods 10, 5–6 (2013).
41. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. G3:Genes|Genomes|Genetics 1, 457–470 (2011).
42. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48, 1279–1283 (2016).
43. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
44. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).

# Supplementary Information

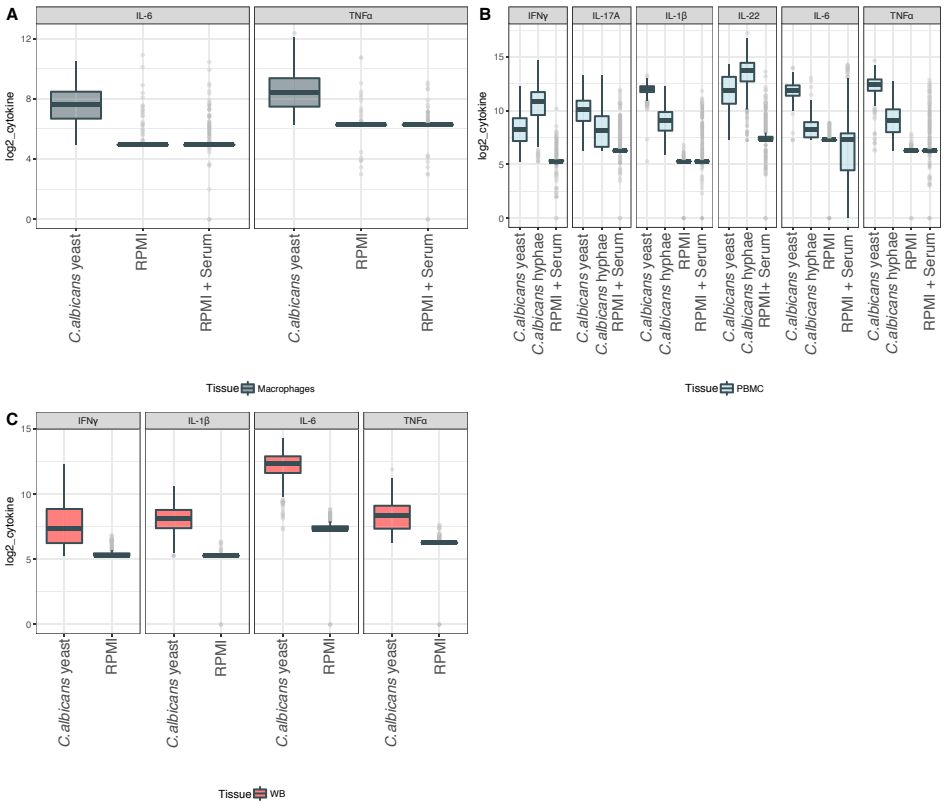



**Figure S1.** Increased cytokine variation upon stimulation with *C. albicans* yeast and hyphae in WB, PBMCs, and monocyte-derived macrophages. Y-axis depicts the log2 transformed cytokine levels. X-axis shows stimulation used: *C. albicans* yeast, *C. albicans* hyphae, or the RPMI medium (with or without serum) used as control. Colors in legend indicate the three different cell systems used. Gray: monocyte-derived macrophages, Light blue: peripheral blood mononuclear cells (PBMCs) and Red: whole blood (WB).

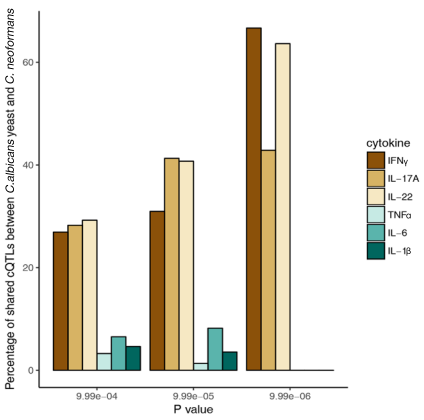



**Figure S2.** Shared genetics between *C. albicans* yeast and *C. neoformans*. (A) Barplot shows a significant enrichment of shared cQTLs that affect T-cell-derived cytokines (IFNγ, IL-17A and IL-22) compared to cQTLs that affect monocyte-derived cytokines at different P-value thresholds: $P < 1x10^{-4}$, $P < 1x10^{-5}$ and $P < 1x10^{-6}$.

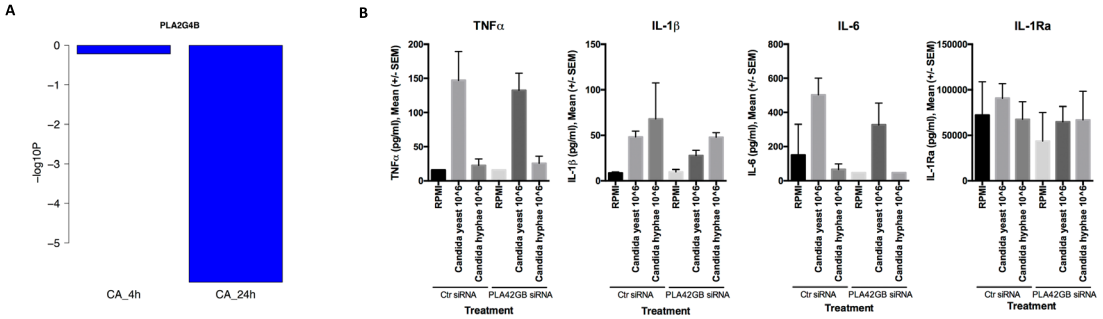



**Figure S3.** *PLA2G4B* locus was prioritized as a plausible candidaemia susceptibility gene based on differential expression and immunological data. (A) Barplot showing the –log10 P values of differential expression of *PLA2G4B* upon stimulation with *C. albicans* in PBMCs compared to the RPMI medium used as control. *PLA2G4B* gene expression was significantly decreased after 24-hour stimulation with *C. albicans* ($P$ $1.02x10^{-13}$). (B) Knockdown efficiency of siRNA experiments of *PLA2G4B* gene upon stimulation of macrophages with *C. albicans* yeast and hyphae. Cytokine production capacity of TNFα, IL-1β, IL-6 and IL-1Ra upon specific *PLA2G4B* siRNA, or control siRNA of *C. albicans* stimulated macrophages. Data from three experiments with at least two donors per experiment was pooled.

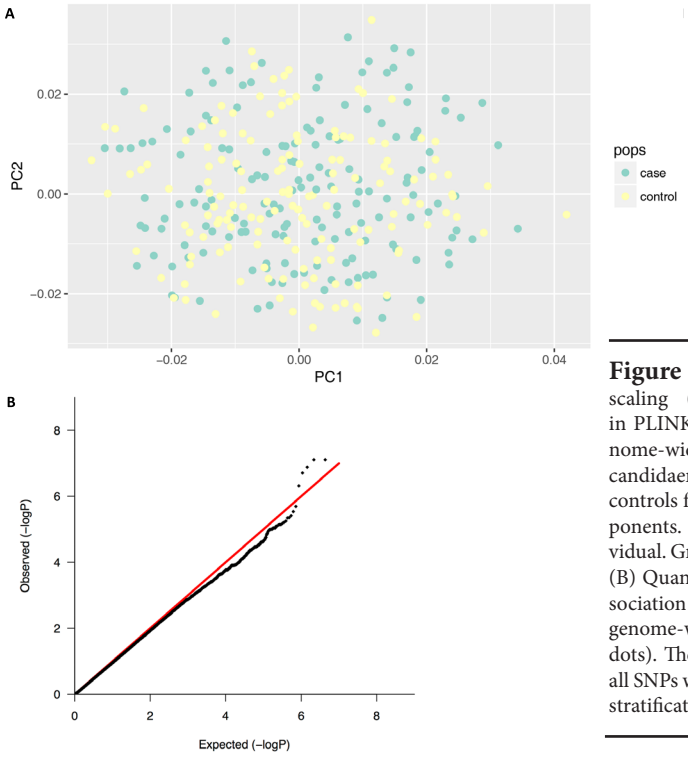



**Figure S4.** (A) Multidimensional scaling (MDS) analysis (performed in PLINK, on the N x N matrix of genome-wide IBS pairwise distance) of candidaemia patients and case-matched controls for the two first principle components. Each dot represents an individual. Green: patients. Yellow: controls. (B) Quantile-Quantile (QQ) plot of association P-values for all SNPs on the genome-wide genotyping array (Black dots). The genetic inflation factor λ for all SNPs was 1, indicating no population stratification.

**Table S1.** Twenty SNPs associated with candidaemia susceptibility (out of seventy-five, $P < 0.05$) modulate multiple cytokines upon stimulation with the two *C. albicans* morphotypes, yeast and hyphae, of any of the three cell systems (WB: whole blood, PBMCs: peripheral blood mononuclear cells, and macrophages).

| Chrom | SNP | Cytokine | Stimulant | Cell system | Time | P value | | OR |
|---|---|---|---|---|---|---|---|---|
| | | | | | | cQTL | Susceptibility | |
| 2 | rs35523523 | TNFα | *C. albicans* hyphae | PBMC | 24 hours | $3.91x10^{-2}$ | $5.17x10^{-5}$ | 2.2 |
| 2 | rs35523523 | TNFα | *C. albicans* yeast | whole blood | 48 hours | $3.27x10^{-2}$ | $5.17x10^{-5}$ | 2.2 |
| 5 | rs4895365 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $2.22x10^{-2}$ | $6.94x10^{-5}$ | 0.46 |
| 5 | rs4895365 | IFNγ | *C. albicans* yeast | PBMC | 7 days | $2.32x10^{-3}$ | $6.94x10^{-5}$ | 0.46 |
| 5 | rs4895365 | IL-22 | *C. albicans* yeast | PBMC | 7 days | $2.89x10^{-3}$ | $6.94x10^{-5}$ | 0.46 |
| 13 | rs11149101 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $4.02x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs11149101 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $4.47x10^{-3}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs11149101 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.19x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs12429614 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $3.95x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs12429614 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $4.49x10^{-3}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs12429614 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.22x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs17069905 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $2.88\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs17069905 | TNFα | *C. albicans* hyphae | PBMC | 24 hours | $4.69x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs17069905 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $2.47x10^{-3}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs17069905 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.32x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs17069905 | TNFα | *C. albicans* yeast | macrophages | 24 hours | $4.80x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs2329096 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $4.01x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs2329096 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $4.47x10^{-3}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs2329096 | IL-6 | C. albicans yeast | macrophages | 24 hours | $4.17x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs4562959 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $2.78x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs4562959 | TNFα | *C. albicans* hyphae | PBMC | 24 hours | $4.52\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs4562959 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $2.46\ x10^{-3}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs4562959 | IL-6 | C. albicans yeast | macrophages | 24 hours | $4.35\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs4562959 | TNFα | *C. albicans* yeast | macrophages | 24 hours | $4.78\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs56160362 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $2.78\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs56160362 | TNFα | *C. albicans* hyphae | PBMC | 24 hours | $4.52\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs56160362 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $2.46\ x10^{-3}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs56160362 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.35\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs56160362 | TNFα | *C. albicans* yeast | macrophages | 24 hours | $4.78\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs56172265 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $3.95\ x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs56172265 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $4.49\ x10^{-3}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs56172265 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.22\ x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs6563046 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $4.94\ x10^{-2}$ | $4.35x10^{-6}$ | 2.12 |
| 13 | rs6563046 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $1.91\ x10^{-3}$ | $4.35x10^{-6}$ | 2.12 |
| 13 | rs7322708 | IL-22 | *C. albicans* hyphae | PBMC | 7 days | $3.43\ x10^{-2}$ | $2.08x10^{-5}$ | 2.34 |
| 13 | rs7322708 | IL-1β | *C. albicans* yeast | whole blood | 48 hours | $3.59\ x10^{-2}$ | $2.08x10^{-5}$ | 2.34 |
| 13 | rs7985222 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $2.78\ x10^{-2}$ | $4.36x10^{-5}$ | 2.01 |
| 13 | rs7985222 | TNFα | *C. albicans* hyphae | PBMC | 24 hours | $4.52\ x10^{-2}$ | $4.36x10^{-5}$ | 2.01 |
| 13 | rs7985222 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $2.46\ x10^{-3}$ | $4.36x10^{-5}$ | 2.01 |
| 13 | rs7985222 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.35\ x10^{-2}$ | $4.36x10^{-5}$ | 2.01 |
| 13 | rs7985222 | TNFα | *C. albicans* yeast | macrophages | 24 hours | $4.78\ x10^{-2}$ | $4.36x10^{-5}$ | 2.01 |
| 13 | rs9318560 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $4.78\ x10^{-2}$ | $4.36x10^{-5}$ | 2.01 |
| 13 | rs9318560 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $2.72\ x10^{-3}$ | $4.36x10^{-5}$ | 2.01 |
| 13 | rs9318560 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.01\ x10^{-2}$ | $4.36x10^{-5}$ | 2.01 |
| 13 | rs9565418 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $2.81\ x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565418 | TNFα | *C. albicans* hyphae | PBMC | 24 hours | $4.57x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565418 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $2.46x10^{-3}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565418 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.34x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565418 | TNFα | *C. albicans* yeast | macrophages | 24 hours | $4.79x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565419 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $2.85x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565419 | TNFα | *C. albicans* hyphae | PBMC | 24 hours | $4.64x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565419 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $2.47x10^{-3}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565419 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.33x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |

| Chrom | SNP | Cytokine | Stimulant | Cell system | Time | P value | | OR |
|---|---|---|---|---|---|---|---|---|
| | | | | | | cQTL | Susceptibility | |
| 13 | rs9565419 | TNFα | *C. albicans* yeast | macrophages | 24 hours | $4.79x10^{-2}$ | $2.96x10^{-5}$ | 2.04 |
| 13 | rs9565420 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $3.99x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs9565420 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $5.31x10^{-3}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs9565420 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $4.52x10^{-2}$ | $2.73x10^{-5}$ | 2.06 |
| 13 | rs9565423 | IL-1β | *C. albicans* hyphae | PBMC | PBMC | $4.86x10^{-2}$ | $6.17x10^{-6}$ | 2.1 |
| 13 | rs9565423 | TNFα | *C. albicans* yeast | PBMC | 24 hours | $1.93x10^{-3}$ | $6.17x10^{-6}$ | 2.1 |
| 18 | rs12955300 | IL-17A | *C. albicans* hyphae | PBMC | 7 days | $3.19x10^{-2}$ | $3.79x10^{-5}$ | 0.48 |
| 18 | rs12955300 | IL-17A | *C. albicans* yeast | PBMC | 7 days | $3.37x10^{-2}$ | $3.79x10^{-5}$ | 0.48 |
| 18 | rs12955300 | IL-6 | *C. albicans* yeast | PBMC | 24 hours | $3.97x10^{-2}$ | $3.79x10^{-5}$ | 0.48 |
| 18 | rs56126175 | IL-17A | *C. albicans* hyphae | PBMC | 7 days | $2.07x10^{-2}$ | $1.23x10^{-5}$ | 0.44 |
| 18 | rs56126175 | TNFα | *C. albicans* yeast | macrophages | 24 hours | $1.83x10^{-2}$ | $1.23x10^{-5}$ | 0.44 |
| 18 | rs56126175 | IFNγ | *C. albicans* yeast | PBMC | 7 days | $3.68x10^{-2}$ | $1.23x10^{-5}$ | 0.44 |
| 18 | rs56126175 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $3.77x10^{-2}$ | $1.23x10^{-5}$ | 0.44 |
| 18 | rs72987756 | IL-17A | *C. albicans* hyphae | PBMC | 7 days | $3.14x10^{-2}$ | $2.65x10^{-5}$ | 0.45 |
| 18 | rs72987756 | TNFα | *C. albicans* yeast | macrophages | 24 hours | $1.57x10^{-2}$ | $2.65x10^{-5}$ | 0.45 |
| 18 | rs72987756 | IL-6 | *C. albicans* yeast | macrophages | 24 hours | $3.29x10^{-2}$ | $2.65x10^{-5}$ | 0.45 |

**Table S2.** Pathway enrichment analysis of genes extracted from a window of 500 Kb upstream and downstream of each of the twenty-one *C. albicans* response cQTLs that are associated with candidaemia susceptibility. The p-values are corrected for multiple testing using the false discovery rate method and are available as q-values. Reactome was used as a database resource of the molecular pathways.

| p-value | q-value | Pathway | Members input overlap |
|---|---|---|---|
| 1.52x10-3 | 2.52x10-3 | Metabolism of lipids and lipoproteins | *TNFAIP8; JMJD7-PLA2G4B; AGPAT1; GGT5; ALOX15B; HSD17B4* |
| 1.81x10-3 | 2.94x10-3 | Synthesis of PA | *JMJD7-PLA2G4B; AGPAT1* |
| 2.25x10-3 | 3.49x10-3 | Glutathione conjugation | *GSTT1; GGT5* |
| 5.42x10-3 | 7.75x10-3 | Arachidonic acid metabolism | *ALOX15B; GGT5* |
| 6.16x10-3 | 8.65x10-3 | Phospholipid metabolism | *JMJD7-PLA2G4B; TNFAIP8; AGPAT1* |

# CHAPTER 4

# Circulatory protein profiles in plasma of candidaemia patients and the contribution of host genetics to their variability

Vasiliki Matzaraki, Kieu Le, Rob Ter Horst, Martin jaeger, Raul Aguirre-Gamboa, Melissa D Johnson, Serena Sanna, Urmo Vosa, Lude Franke, Alexandra Zhernakova, Jingyuan Fu, Sebo Withoff, Iris Jonkers, Yang Li, Leo A.B. Joosten, Mihai G Netea, Cisca Wijmenga, Vinod Kumar

# CHAPTER 4

# Circulatory protein profiles in plasma of candidaemia patients and the contribution of host genetics to their variability

Vasiliki Matzaraki[1], Kieu Le[1,*], Rob Ter Horst[2,*], Martin jaeger[2], Raul Aguirre-Gamboa[1], Melissa D Johnson[3], Serena Sanna[1], Urmo Vosa[1], Lude Franke[1], Alexandra Zhernakova[1], Jingyuan Fu[1,4], Sebo Withoff[1], Iris Jonkers[1], Yang Li[1], Leo A.B. Joosten[2], Mihai G Netea[2,5], Cisca Wijmenga[1,6], Vinod Kumar[1,2]

[1]University of Groningen, University Medical Center Groningen, Department of Genetics, 9713 GZ Groningen, The Netherlands.
[2]Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands.
[3]Duke University Medical Center, Durham, North Carolina, United States of America.
[4]University of Groningen, University Medical Center Groningen, Department of Pediatrics, 9713 GZ Groningen, The Netherlands.
[5]Department for Genomics and Immunoregulation, Life and Medical Sciences Institute (LIMES), University of Bonn, 53115 Bonn, Germany.
[6]Department of Immunology, K.G. Jebsen Coeliac Disease Research Centre, University of Oslo, 0424 Oslo, Norway.

**Abstract**

Circulatory inflammatory proteins, such as cytokines and chemokines, play a significant role in anti-*Candida* host immune defence. However, little is known about the genetic variation that contributes to the variability of inflammatory responses in response to *C. albicans*. To systematically characterize inflammatory responses in *Candida* infection, we profiled 92 circulatory inflammatory proteins in 42 candidaemia patients. Given the significant differences in inflammatory protein profiles between patient and healthy individuals, we correlated genome-wide single nucleotide polymorphism (SNP) genotypes with protein abundance (QTLs) produced by peripheral blood mononuclear cells, stimulated with *C. albicans* yeast, from 436 individuals of European origin from the 500 Functional Genom-

ics (500FG) cohort in the Human Functional Genomics Project and Lifelines Deep cohort. We identified 10 genome-wide significant protein-QTLs modulated CCL4, VEGF-A, IL-8, CXCL9, MCP-1, MCP-2 and MCP-3 in response to *C. albicans*. Furthermore, we investigated whether differences in susceptibility and survival of candidaemia patients can be explained by modulating levels of inflammatory proteins. Our genetic analysis suggested that there is a distinct genetic contribution between inflammatory responses to *Candida* infection, susceptibility, and survival, indicating a different biology underlying these phenotypes that could provide new therapeutic opportunities.

**Keywords:** inflammatory proteins, protein-QTLs, *C. albicans*, candidaemia, susceptibility, survival

## Introduction

*Candida* species are by far the most common fungal pathogens that cause both invasive and mucosal fungal infections. They have been described as the fourth most common cause of nosocomial bloodstream infection in the United States[1,2]. Invasive candidiasis causes more than 250,000 new systemic infections on a yearly basis and leads to more than 50,000 deaths[3]. Most humans are colonized with *C. albicans* shortly after birth, which remains as part of a normal human's microbiota. Infection occurs only if the epithelial barrier function is impaired and/or there are microbiome imbalances and/or the host immune system is compromised. Under these conditions, *Candida* can invade tissue and reach blood circulation. The bloodstream carries *Candida* to almost all vital organs, leading to systemic infections and, eventually, to organ failure followed by death.

Protective immunity to *Candida* involves both innate and adaptive cellular and humoral responses[4,5]. Cytokines and chemokines are a group of low molecular weight proteins that contribute significantly to anti-*Candida* host immune defence by acting as mediators between immune and non-immune cells, by enhancing the antifungal activity of immune cells and by attracting inflammatory immune cells to the site of infection. Previous studies have shown the capacity of *C. albicans* to induce production of various cytokines and chemokines[6–8]. Some of these proteins in the circulation have also been extensively explored as disease biomarkers to determine onset, progression, patient susceptibility, or predict efficacy of a treatment[9–11]. However, proteins are inherently influenced by genetic and non-genetic factors. Therefore, the identification of these factors would help

to stratify patients based on their risk, and those at high risk would benefit most by prophylactic antifungal treatment or adjunctive immunotherapy.

By studying genetics of only six different cytokines in the context of *Candida*-stimulation, we have shown that SNPs affecting cytokine responses are associated to susceptibility to candidaemia[12–14], suggesting that modulation of cytokines can determine disease susceptibility. Of note, among candidaemia patients with a combination of several well-described risk factors, the disease prognosis is generally poor and survival rates differ greatly among them, indicating that genetic variation determines patient survival as well[15]. Given that genetic factors regulate the excess levels of proteins in circulation or dysregulated production, which can, subsequently, determine susceptibility and patient survival, it is important to assess the impact of these genetic factors on a wide range of inflammatory proteins in circulation during *C. albicans* infections. In the current study, therefore, we hypothesized that modulation of circulatory inflammatory proteins contributes to susceptibility to candidaemia and patient survival (Fig. 1A). However, all previous studies with *C. albicans* as stimulant studied a narrow spectrum of inflammatory proteins, such as cytokines and chemokines[6–8], and a systematic study of inflammatory proteins released in the blood circulation upon *C. albicans* and their impact on susceptibility and patient survival is lacking.

Thus, the aims of the present study were to identify (i) the abundance of differentially regulated plasma proteins in circulation of patients (Fig. 1B) (ii) (non)-genetic factors that may influence the abundance of inflammatory proteins in circulation (Fig. 1C and D) and, (iii) finally, to investigate the impact of these genetic loci on susceptibility to candidaemia and patient survival. For this, we profiled 92 inflammatory proteins in the plasma of a candidaemia patient cohort and of two independent population-based cohorts, the 500 Functional Genomics (500FG) cohort within the Human Functional Genomics Project (http://www.human-function-algenomics.org) and Lifelines DEEP cohort (https://www.lifelines.nl/researcher/biobank-lifelines/additional-studies/lifelines-deep) using the Olink inflammatory array (http://www.olink.com).

We observed a significantly different circulatory protein profile in plasma from candidaemia patients compared to healthy individuals. In addition, we identified 10 independent novel protein quantitative trait loci (pQTLs, $P < 5x10^{-8}$). Of note, pQTLs showed a poor enrichment for genetic variants associated with candidaemia susceptibility and patient survival. This finding indicates a distinct genetic contribution in candidaemia susceptibility, patient survival and inflammatory responses in *Candida* infection. We finally investigated whether pQTLs also contribute to other complex infectious diseases.

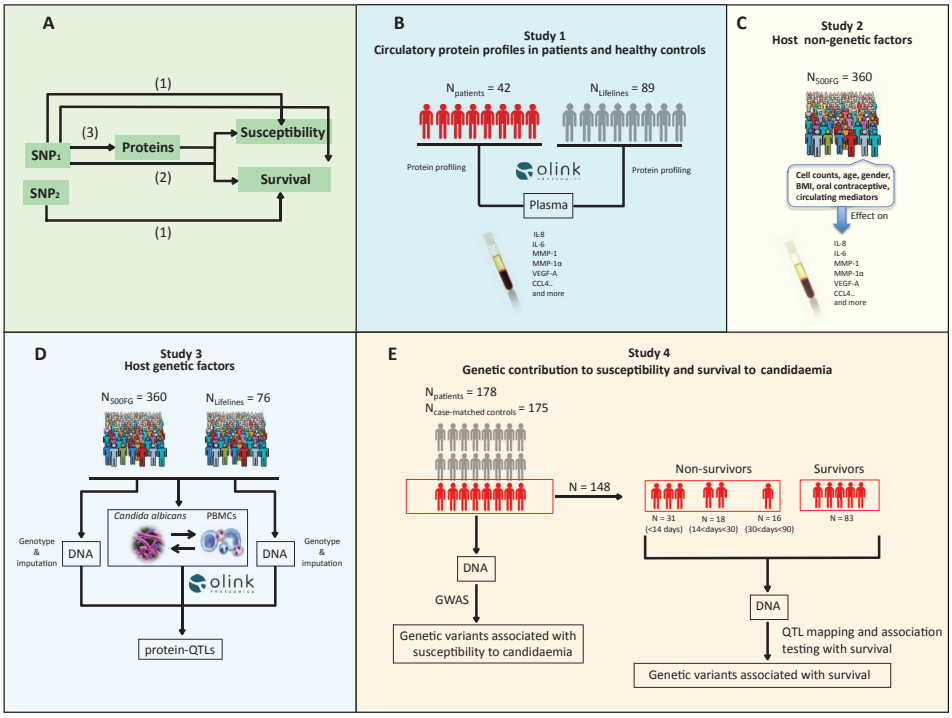



**Fig. 1.** Overview of our hypotheses and studies. (A) Three scenarios can explain susceptibility to candidaemia and patient survival: (1) Distinct genetic variation ($SNP_1$ or $SNP_2$) determines susceptibility to candidaemia or patient survival, (2) same genetic variation contribute to the two phenotypes (susceptibility or survival) and (3) the two phenotypes can be determined by modulating the levels of inflammatory proteins (same or different pQTLs) in blood circulation. (B) By obtaining the protein profiling of plasma proteins in candidaemia patients (n = 42) and healthy individuals from Lifelines cohort (n = 89), we compared the abundance of differentially regulated plasma proteins in blood circulation between patients and healthy subjects. (C) We also studied the effect of non-genetic factors available for 360 individuals from 500FG cohort on the regulation of various inflammatory proteins. (D) In addition to non-genetic factors, we studied the effect of host genetics on the regulation of inflammatory responses by mapping pQTLs in a joint analysis of 500FG (n = 360) and Lifelines Deep (n = 76) cohorts. For pQTL mapping, we profiled the proteins released from *C. albicans*-stimulated PBMCs isolated from healthy individuals and obtained the imputed genotypes of the studied individuals. Lastly, (E) we investigated whether pQTLs determine susceptibility to candidaemia and patient survival. For this, we used data from our previous GWAS of candidaemia patients (n = 178) and case-matched controls (n=175) on candidaemia susceptibility (*chapter 3*) and, in the current study, we performed a QTL mapping of variants associated with survival. In addition, we tested for association with 14-, 30- and 90-day survival by following up a subgroup of candidaemia patients (n = 148), who passed away (1) within 14 days, (2) between 14 and 30 days, and (3) between 30 and 90 days. The rest of the patients (n = 83) survived longer than 90 days (survivors). Protein profiling of candidaemia and population-based cohorts were done using the same inflammatory array from OLINK technology. Stimulations of PBMCs with *C. albicans* were performed for 24 hours using *C. albicans* yeast. GWAS: genome-wide association study; QTL: quantitative trait loci; PBMCs: peripheral blood mononuclear cells.

## Results

### Overview of inflammatory responses in candidaemia patients

To systematically study the inflammatory responses in candidaemia patients compared to healthy individuals, we first assessed the abundance of 92 inflammatory proteins in the plasma of 42 candidaemia patients and 89 healthy individuals from the Lifelines DEEP cohort using the OLINK inflammatory array (Fig. 1B). Out of 92 proteins, 64 proteins were included for the differential abundance analysis based on their detectability in at least 90% of the samples tested. A complete list of the proteins measured is provided in Table S1. Significant differences in the levels of the majority of inflammatory proteins were observed in patient plasma samples compared to baseline healthy plasma samples (56 out of 64 proteins with $P < 0.05$, Figure S1), indicating a high inflammation status in patients compared to healthy controls. Of the significantly differentially expressed proteins, fifteen inflammatory proteins showed more than 1.5 fold differences (Table S2).

To explore further the co-regulation patterns of plasma proteins in patients compared to healthy individuals, we performed unsupervised clustering of protein responses that were detected in at least 90% of the samples and were expressed in both patients and controls ($N_{proteins} = 64$). Clustering indicated that there is a clear distinction in inflammatory responses between patients and healthy controls. Ten patients formed a distinct cluster from the rest of the patients and present a similar clustering pattern as healthy controls, with the exception of MMP-1 expression. MMP1- expression seems to be very strong in all patients compared to controls, suggesting that MMP-1 plays an important role in inflammation in patients (Figure S2).

### Overview of inflammatory protein profiles in PBMCs from healthy individuals in response to *C. albicans* stimulation

By comparing circulatory protein profiles between candidaemia patients and healthy controls, we identified significant differences in inflammatory protein profiles (Figure S1). We, therefore, tested whether some of these proteins are produced by PBMCs in response to *C. albicans* stimulation (Fig. 1D). We obtained PBMCs from 360 individuals of European origin from the 500FG and profiled 92 inflammatory proteins in response to *C. albicans* yeast, of which 32 were detected in at least 90% of the samples (Table S3). Firstly, we compared the abundance of the 32 proteins between stimulated and un-stimulated (using RPMI medium) conditions and observed increased inter-individual variation in inflammatory proteins upon stimulation compared to samples stimulated with RPMI medium (Figure S3). This observation suggested the role of different factors in regulating

the abundance of these proteins.

We could identify up-regulation of 29 proteins out of 32 proteins measured in at least 90% of PBMC samples that show more than 1.5 fold change compared to RPMI stimulated samples (Table S4). In candidaemia patients, we identified 64 proteins measured in at least 90% of the samples, of which 15 showed 1.5 fold difference compared to baseline healthy samples (Lifelines cohort). These observations suggest that, in addition to PBMCs, many other cell types may contribute to the abundance of different proteins in circulation. Next, we compared the correlation structure between the circulatory proteins released from PBMCs by performing an unsupervised clustering of the protein responses. Clustering revealed three distinct clusters (Fig. 2), in which a large cluster consisted of 27 proteins (out

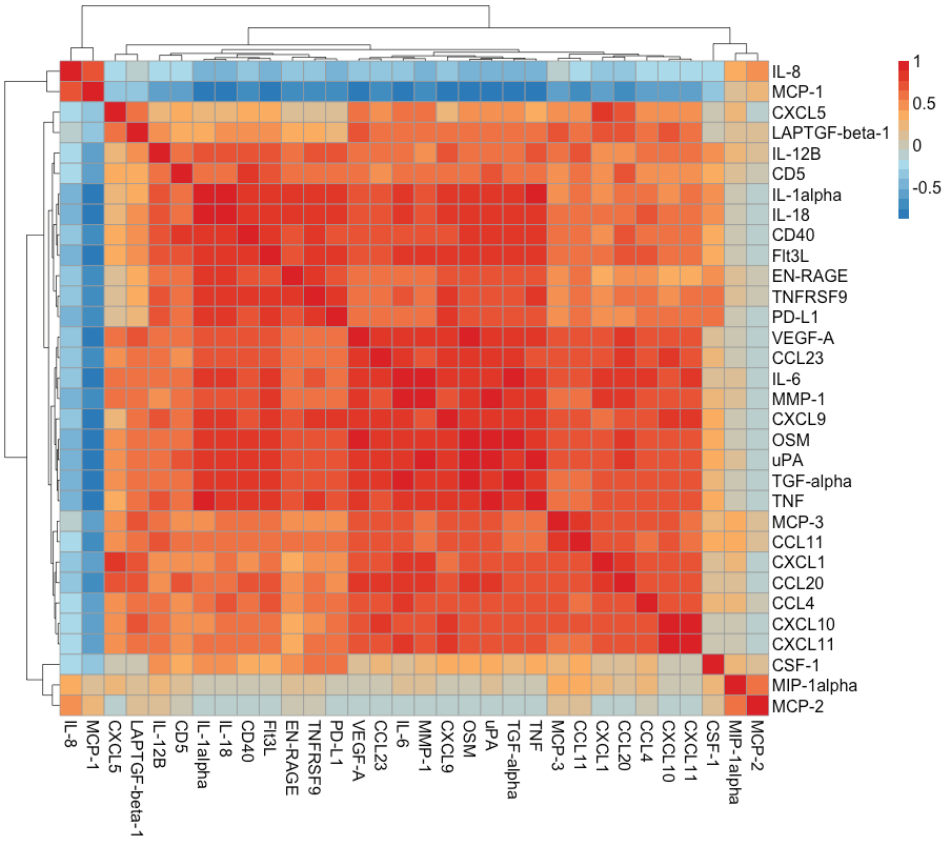



**Fig. 2.** Clustering of inflammatory responses induced in *Candida*-stimulated PBMCs revealed three distinct clusters, with the majority of them organized in a common cluster. The correlation structure of thirty-two proteins, which were detected in at least 90% of the samples, is shown. Unsupervised hierarchical clustering was performed using Spearman correlation as the measure of similarity. The red color depicts the strong positive correlation whereas blue color indicates the strong negative correlation. PBMCs were obtained from 360 individuals from the 500FG cohort and were stimulated with *C. albicans* yeast for 24 hours.

of 32 that were detected in at least 90% of the samples), and two small clusters with only two (IL-8 and MCP-1) and three (CSF-1, MCP-2, and MIP-1 alpha) proteins respectively. To explore if any of these clusters are different in patients, we performed unsupervised clustering of inflammatory proteins released from PBMCs and patients. We observed different clustering patterns of protein responses between PBMC samples and patients (with the exception of one patient) (Figure S4), suggesting the presence of additional factors contributing to the inflammatory responses in patients. MMP-1 expression seems to be stronger in patients compared to 500FG samples, providing more evidence that MMP-1 expression is a distinctive protein signature for patients compared to controls (as mentioned above) and to *Candida*-stimulated PBMCs.

**Minimal effect of non-genetic host factors and cell counts on inflammatory proteins in response to *C. albicans***

Given the increased inter-individual variation in inflammatory proteins upon *Candida* stimulation (Figure S3), we next aimed to identify whether host non-genetic factors determine these person-to-person differences in inflammatory responses. It is possible that different cell-count proportions in individuals can explain this inter-individual variation. To test whether cell-counts influence the levels of inflammatory proteins, in addition to protein data measured in *Candida*-stimulated PBMCs isolated from the 500FG cohort (n = 360), we used the FACS measurements that were previously performed in the same cohort, where different immune cell populations were counted in detail together with non-genetic host factors[16,17]. We analyzed the correlation structure between cell counts ($CD14^+$ monocytes, $CD4^+$ T cells, $CD3^+CD56^-$ T cells, lymphocytes, $CD3^-CD56^+$ NK cells, $CD19^+$ B cells and $CD8^+$ T cells) and protein measurements from *Candida*-stimulated PBMCs. We observed weak correlations between cell counts and protein levels, suggesting a minor effect of cell-count differences on protein production capacity (mean correlation coefficient across $CD14^+$ monocytes = 0.234, $CD4^+$ T cells = -0.056, $CD3^+$ $CD56^-$ T cells = -0.038, lymphocytes = -0.025, $CD3^-$ $CD56^+$ NK cell = 0.01, $CD19^+$ B cells = 0.019 and $CD8^+$ T cells = 0.011) (Figure S5A).

It has been well established that age and gender influence immune responses[18]. Also, we have previously reported a strong impact of age and gender on the production of six different cytokines (IL-1β, TNFα, IL-6, IFNγ, IL-22 and IL-17)[16]. To systematically investigate the impact of non-genetic factors, including age and gender, on the 32 inflammatory proteins released from PBMCs, we correlated age, gender, BMI, oral contraceptive use, and circulating mediators (resistin, adiponectin, alpha-1 Antitrypsin, leptin and CRP) with our proteins. BMI, oral contraceptive use and circulating mediators had no detectable effect on *in vitro*

protein production in PBMCs (Table S5 and S6). In contrast, three of the inflammatory proteins (MMP-1 and CXCL5 and CCL23) were significantly correlated with age ($P < 0.05$) (Figure S5B and Table S5). In addition, TNFRSF9, PD-L1, IL-1aplha, IL-18, IL-12B, EN-RAGE, CXCL9 and CSF-1 showed a significant (positive) correlation with gender ($P < 0.05$) (Figure S5B and Table S5).

**Identifying genetic variation affecting inflammatory proteins in response to *C. albicans***

Next, we investigated whether host genetic variation affects the inter-individual differences in inflammatory responses to *C. albicans* stimulation. For this, we used the genome-wide SNP genotype data and protein measurements of *Candida*-stimulated PBMCs isolated from two population-based cohorts consisting of 360 individuals from the 500FG cohort and 76 individuals from the Lifelines DEEP cohort. Upon quality control and intersection of the proteins from the two cohorts, we obtained a total of 32 inflammatory proteins (Figure S6 and S7 and Table S3 and S7). For pQTL mapping, we selected SNPs that showed a minor allele frequency (MAF) ≥ 1% and passed other quality filters (see Materials and Methods). Using the protein and genotype data, we mapped protein-QTLs (pQTLs) in the two cohorts and performed a joint analysis by combining the two cohorts (n = 436).

In detail, raw protein levels were log2-transformed and then mapped to genotype data using a linear model with age and gender as covariates. Given that a strong influence of age and sex on immunological traits has been previously reported, we used age and gender as covariates on our model[16,19,20]. Our joint analysis revealed 10 independent *trans*-pQTLs with MAF > 1% that reached genome-wide significance level ($P < 5x10^{-8}$ and FDR 0.10) (Table 1, Fig. 3 and Figure S8). These include one independent pQTL for CCL4, one for VEGF-A, two for IL-8, two for CXCL9, one for MCP-1, two for MCP-2 and one for MCP-3 (Fig. 3A-G).

**Prioritizing *cis*-genes from genome-wide significant pQTLs indicates those involved in hemostasis and lipid metabolism as potential causal genes**

We hypothesized that genes from the genome-wide significant pQTL loci that are differentially expressed in PBMCs in response to *C. albicans* are potential causal genes. To investigate this, we tested the expression levels of all genes located within a 500 kb *cis*-window of the pQTLs in PBMCs stimulated with *C. albicans* at 4 and 24 hours. We identified Wnt family member 5A (*WNT5A*) gene as differentially expressed at both 4 and 24 hours. Single-stranded DNA binding protein 2 (*SSBP2*), family with sequence similarity 151 member B (*FAM151B*) and caspase 7 (*CASP7*) genes were differentially expressed at 24 hours. Of those, *WNT5A* gene

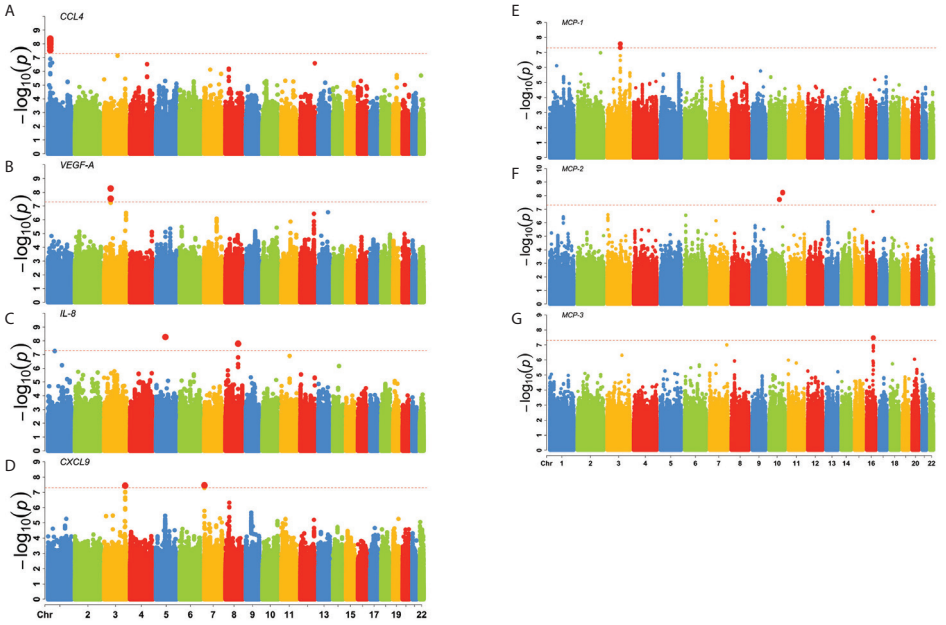



**Fig. 3.** Genome-wide pQTL mapping identified ten *C. albicans* yeast-response pQTLs. Manhattan plot showing the genome-wide QTL mapping results for *C. albicans*-induced (A) CCL4, (B) VEGF-A, (C) IL-8, (D) CXCL9, (E) MCP-1, (F) MCP-2 and (G) MCP-3 levels. The y-axis represents the –log10P values of pQTLs. Their chromosomal positions are shown on the x axis. The horizontal red dashed line represents the genome-wide significance threshold for association ($P < 5x10^{-8}$).

**Table 1.** Genome-wide significant protein-QTL loci that were identified in the joint analysis.

| Chr | SNP | BP | Allele1 | Allele2 | MAF | Protein | Z score | P value | *cis*-genes |
|---|---|---|---|---|---|---|---|---|---|
| 1 | rs2501301 | 22344027 | T | C | 0.05 | CCL4 | 5.88 | $4.17x10^{-9}$ | *CELA3B*[a], *CDC42-IT1*[b], *HSPG2*[c] *ABHD10*[d], *CD96*[d], *ZBED2*[c] |
| 3 | rs1398749 | 111456265 | A | G | 0.09 | MCP-1 | -5.56 | $2.70x10^{-8}$ | *WNT5A*[c,e] |
| 3 | rs358011 | 55125391 | T | C | 0.24 | VEGF-A | 5.84 | $5.23x10^{-9}$ | *EPHB3*[c] |
| 3 | rs6771739 | 184506578 | T | C | 0.43 | CXCL9 | 5.51 | $3.69x10^{-8}$ | *CTC-281B15.1*[b], *SSBP2*[c], |
| 5 | rs392422 | 80332413 | A | G | 0.05 | IL-8 | -5.84 | $5.34x10^{-9}$ | *FAM151B*[c] |
| 7 | rs34774255 | 2747852 | T | C | 0.38 | CXCL9 | 5.52 | $3.50x10^{-8}$ | *AMZ1*[f], *CARD11*[f] |
| 8 | rs1699130 | 107338659 | C | G | 0.07 | IL-8 | -5.65 | $1.64x10^{-8}$ | *OXR1*[f] |
| 10 | rs2488633 | 86351356 | A | G | 0.02 | MCP-2 | -5.62 | $1.93x10^{-8}$ | *CDHR1*[c] |
| 10 | rs1572285 | 115907856 | T | C | 0.12 | MCP-2 | -5.83 | $5.70x10^{-9}$ | *CASP7*[c] |
| 16 | rs247426 | 75438172 | T | C | 0.05 | MCP-3 | 5.52 | $3.36x19^{-8}$ | *ZNRF1*[e], *ZNRF1*[c] |

[a]Proxy SNP (LD >= 0.8) is a missense variant. [b]pQTL or proxies (LD >= 0.8) shows a *cis*-eQTL effect upon *C. albicans* stimulation in PBMCs ($P < 0.05$). [c]Gene is differentially expressed upon *C. albicans* stimulation at 24 hours. [d]Proxy SNP shows an eQTL effect in blood. [e]Gene is differentially expressed upon *C. albicans* stimulation at 4 hours. [f]Gene is in close proximity to pQTL.
MAF: Minor allele frequency in the 500FG and Lifelines DEEP cohorts, Chr: chromosome, BP: base-pairs

belongs to the WNT gene family that signals through both the canonical and non-canonical WNT pathways. Of note, the Wnt/β-Catenin signaling pathway exerts immunomodulatory functions during inflammation and infection[21,22]. The other prioritized gene, *CASP7*, a member of the cysteine-aspartic acid protease (caspase) family, is involved in the activation cascade of caspases that are critical molecules in apoptosis, necrosis, and inflammation[23].

In addition, to identify causal genes, we made use of publicly available expression-QTL datasets (eQTLs) from healthy blood donor samples[24,25]. We identified a proxy SNP rs4682062 in strong linkage disequilibrium (LD) with rs1398749 on chromosome 3 ($r^2 > 0.8$; using the European population as a reference) to influence the expression of two genes, the abhydrolase domain containing 10 (*ABHD10, P* $1.58x10^{-27}$) and *CD96* (*P* $1.50x10^{-5}$) genes, in whole blood. *CD96* gene encodes a protein (also known as TACTILE) that belongs to immunoglobulin superfamily and negatively controls cytokine production by natural killer (NK) cells[26]. The *ABDH10* gene encodes for an enzymatic degrader of mycophenolic acid acyl-glucoronide[27]. In general, the mammalian alpha beta hydrolase domain (ABHD) proteins have been recently described as novel potential regulators of lipid metabolism and signal transduction[27]. Given that eQTL effects may act in a context-specific manner[28], we also mapped *Candida*-response eQTLs. Three SNPs in high linkage disequilibrium (LD ≥ 0.8) with pQTL rs2501301 on chromosome 1, influence the expression of a non-coding RNA, *CDC42-IT1*, and pQTL rs392422 on chromosome 5 influences another non-coding RNA, *CTC-281B15.1* in response to *C. albicans* (Table 1).

Lastly, pathway enrichment analysis of *Candida*-induced pQTLs with a *P* value < $9.99x10^{-4}$ showed an over-representation of genes involved in hemostasis (*P* $8.99x10^{-5}$) and metabolism of lipids and lipoproteins (*P* $2.47x10^{-3}$) (Figure S9), indicating the importance of hemostasis and lipid metabolism in anti-*Candida* inflammatory responses.

**Pleiotropy of loci affecting protein levels**

To test the presence of potential pleiotropic effects between the GWAS pQTLs identified in the joint analysis and measured proteins, we performed an unsupervised clustering of pQTLs based on the negative log10 of the P values. For this analysis, we extracted all possible associations of genome-wide significant pQTLs with the 32 proteins in the same allelic direction between the two cohorts and P value of heterogeneity ≥ 0.05. We observed that all genome-wide significant loci influence multiple distinct proteins (Fig. 4A), indicating the pleiotropic effects of the pQTLs on the proteins. To account for multiple testing, we next inspected all genome-wide significant *trans*-pQTLs that had at least two associations with dis-

tinct proteins at $P < 0.05/(10*32) = 1.56x10^{-4}$ (Table S8). This cut-off represents a conservative approach to correct for multiple testing for all identified GWAS SNPs (n = 10) with all protein traits (n = 32). SNP rs248863 was removed as it influences only MCP-2 at P threshold of $1.56x10^{-4}$. We observed a locus at intergenic variant rs6771739 on chromosome 3 and a second locus at intronic variant rs34774255 on chromosome 7 that influences 17 and 11 distinct proteins respectively at $P < 1.95x10^{-4}$ in the same direction (Fig. 4B to C, and Figure S10). Of note, both loci influence the levels of CXCL9 in *trans*, indicating that this chemokine may have a key regulatory role in inflammation in response to *C. albicans* infection.

Moreover, two loci at intergenic variant rs358011 (Fig. 4D) and intronic variant rs247426 (Fig. 4E) on chromosomes 3 and 16 respectively influence six distinct proteins each in the same direction. The variant rs358011 influences the inflammatory protein VEGF-A and the rs247426 influences MCP-3 in *trans* at a genome-wide significance level (Table 1, Figure S10). These pleiotropic effects suggest that MCP-3 and VEGF-A may have important roles in inflammation induced by *C. albicans*. Lastly, unsupervised clustering revealed that different loci influence same proteins in opposite direction (Figure S10). For instance, the locus at intronic variants rs1699130 on chromosome 8 increases the expression of MCP-1 and IL-8 and two loci at intronic variants rs1398749 and rs392422 on chromosomes 3 and 5 respectively show an opposite effect. In addition, the locus at intronic variants rs247426 on chromosome 16 increases the expression of MCP-2 and MCP-3 and a different locus at rs1572285 on chromosome 10 decreases their expression. However, more studies are needed to explain the mechanism by which these opposite effects occur.

## Contribution of pQTLs that influence circulatory inflammatory proteins to susceptibility and survival to candidaemia

Next, we aimed to investigate whether genetic variation that influences inflammatory proteins (pQTLs) in blood circulation may determine susceptibility and survival to candidaemia. We previously performed the first GWAS on the largest candidaemia cohort to date to identify genetic variants that determine susceptibility to candidaemia (*chapter 3*). The candidaemia cohort consisted of 178 patients and 175 case-matched controls. To investigate the contribution of pQTLs to candidaemia susceptibility, we performed enrichment analysis of pQTLs for susceptibility variants. For this, we grouped the pQTLs identified in the joint analysis at four different P thresholds ($5x10^{-4}$, $5x10^{-5}$, $5x10^{-6}$, and $5x10^{-7}$) and extracted the P values of association to candidaemia susceptibility for each group of pQTLs. We observed that there is a poor enrichment of pQTLs in susceptibility variants at all four different P thresholds. (Fig. 5A to D). These results suggest that there is a dis-

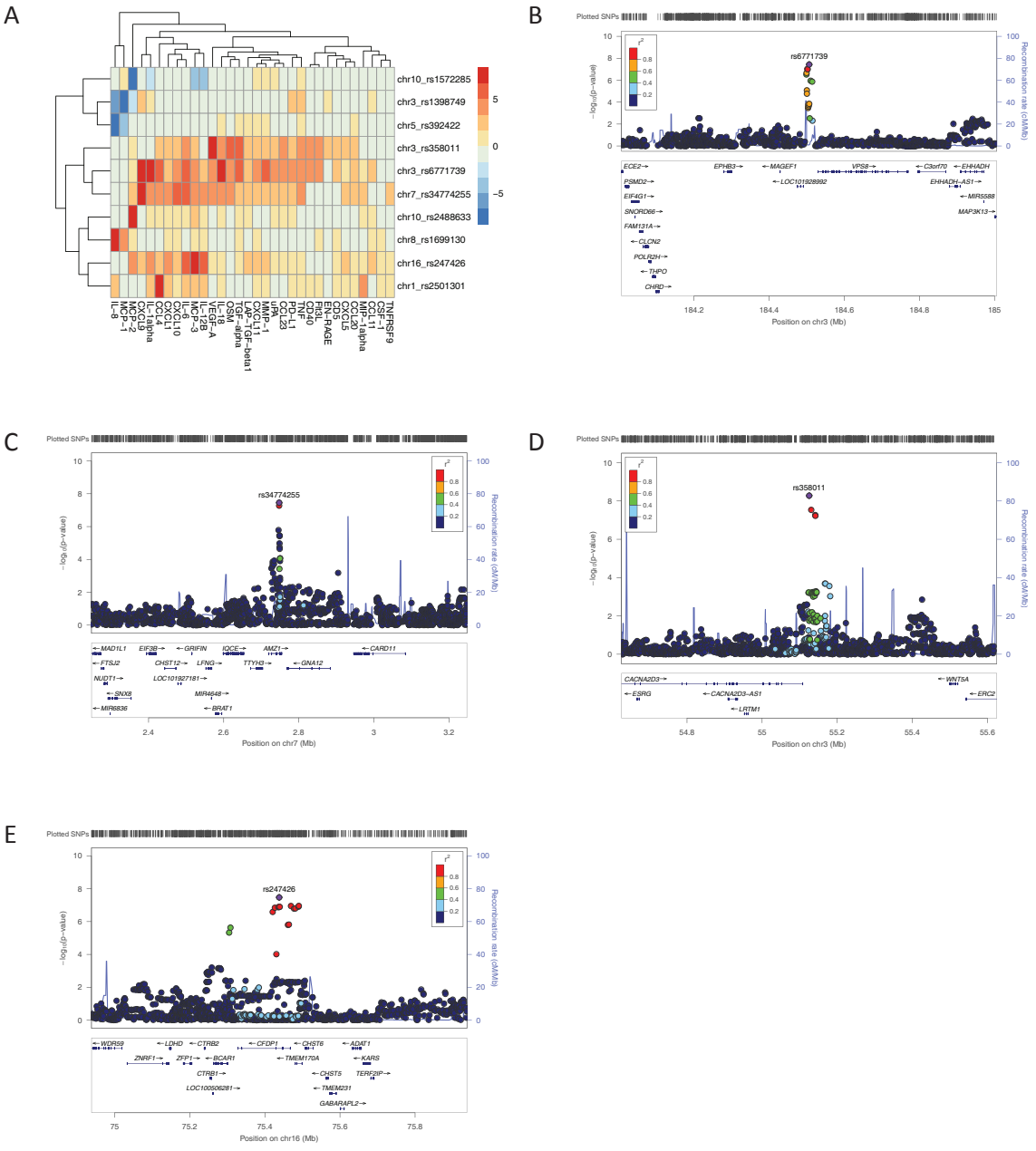



**Fig. 4.** Potential pleiotropy between genome-wide significant pQTLs and measured inflammatory proteins. Heatmap (A) shows the –log10(P) of all possible associations of the genome-wide significant pQTLs and the 32 measured inflammatory proteins, which show the same allelic directions between 500FG and Lifelines Deep cohorts, and $P_{heterogeneity} \geq 0.05$. Regional association plots showing the –log10P for SNPs centered on the most strongly associated signal (purple diamond) at (B) rs6771739 associated with CXCL9 levels, (C) rs34774255 associated with CXCL9 levels, (D) rs358011 associated with VEGF-A and (E) rs247426 associated with MCP-3 levels upon *C. albicans* yeast stimulation of PBMCs.

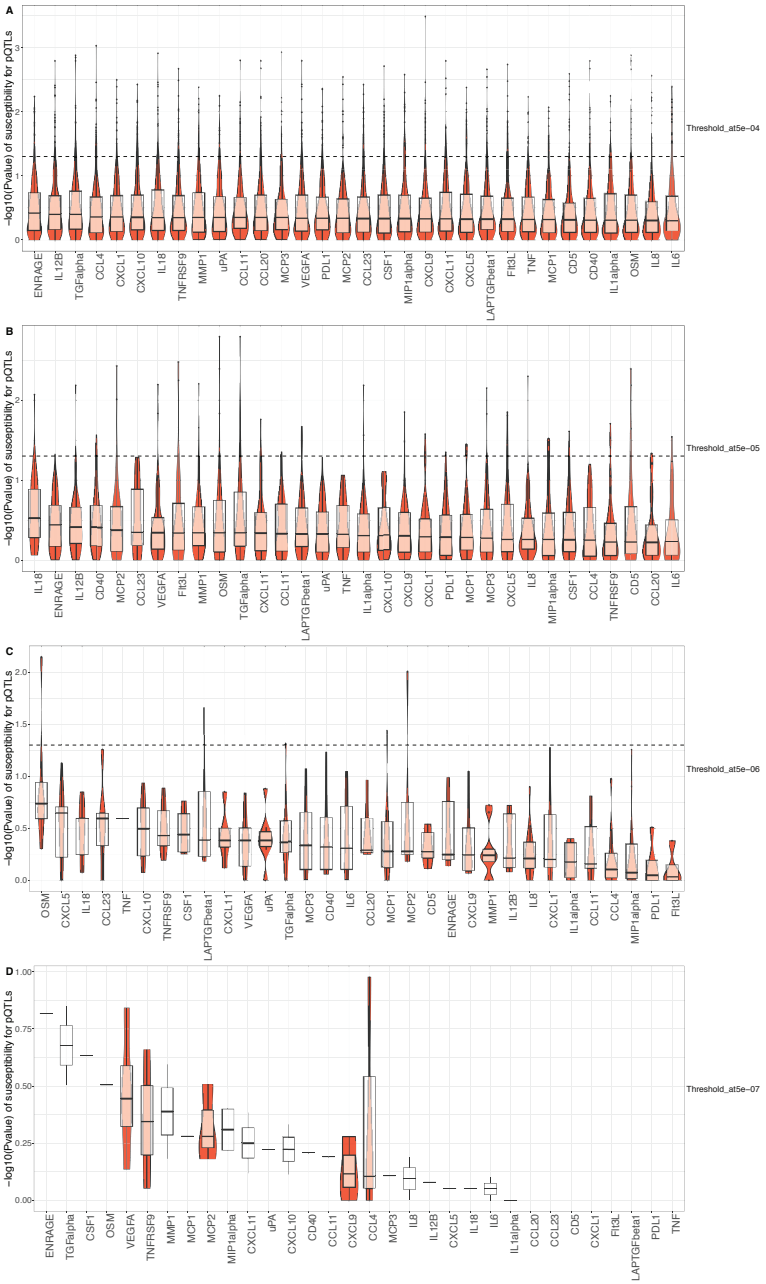



**Fig. 5.** A distinct genetic contribution between candidaemia susceptibility and inflammatory protein responses upon *Candida* stimulation was observed. pQTLs identified in the joint analysis between 500FG and Lifelines DEEP cohorts were grouped at four different P thresholds($5x10^{-4}$, $5x10^{-5}$, $5x10^{-6}$ and $5x10^{-7}$) and P values of association with susceptibility to candidaemia were extracted and plotted for each group (y axis). Axis x shows the 32 inflammatory proteins measured in PBMCs in response to *Candida* stimulation. Violin plots show the distribution of P values of association with susceptibility for each different threshold for pQTL effect. The black line within the box plots indicates the median and thdashed black line represents a P value of significance 0.05. pQTLs above the dashed line were considered to be significantly associated with candidaemia.

tinct genetic contribution between candidaemia susceptibility and inflammatory protein responses upon *Candida* stimulation.

In addition, to determine genetic variants that are associated with survival, we performed a within-case QTL mapping on a genome-wide scale by assigning the

actual day of survival as quantitative trait using a cohort of 148 candidaemia patients, which was followed prospectively for up to 90 days and mortality dates were recorded. Out of 148 patients, 65 did not survive (non-survivors) within a period of 90 days and 83 patients survived longer than 90 days (survivors). For QTL mapping, we used the genotype and survival data of only 65 non-survivors, of which 31 passed away within 14 days, 18 passed away between 14 and 30 days and 16 passed away between 30 and 90 days, as described in Materials and Methods. This analysis showed a common, intergenic genetic variant ($MAF_{1000G}$ = 0.2), rs12565126, on chromosome 1 that reached genome-wide significance (*P* $1.8x10^{-10}$) (Fig. 6A and B). To examine the effect of SNP genotypes of rs12565126 on survival, we compared differences between SNP genotype strata using a log-rank test by stratifying the 65 non-survivors and 83 survivors by SNP genotype. No significant difference in survival was observed between the different genotype strata. However, Kaplan-Meier plot showed that individuals carrying the T allele tend to have an increased probability to survive longer compared to those homozygotes for CC allele (Fig. 6C).

Furthermore, to test whether the genome-wide significant rs12565126 is associated with early or late survival (up to 90 days), we performed a case/control association analysis with 14-, 30- and 90-day survival on a genome-wide scale. To test for association (1) with 14-day survival, we used 31 non-survivors that passed away within 14 days and 83 survivors, with (2) 30-day survival, we used 49 non-survivors that passed away within 30 days and 83 survivors and with (3) 90-day survival, we used 65 non-survivors that passed away within 90 days and 83 survivors. We run an additive model using gender and the first two principal components as covariates. These analyses showed that rs12565126 was significantly associated with 14-day survival (*P* 0.03, OR 6.9) and 30-day survival (*P* 0.02, OR 3.6). We did not observe a significant association with 90-day survival at a P threshold for significance < 0.05 (*P* = 0.54).

Lastly, to investigate the contribution of pQTLs to survival, we performed enrichment analysis of pQTLs for variants associated with survival as identified in the quantitative analysis. For this, we followed the same approach as in enrichment analysis of pQTLs for candidaemia susceptibility, and we grouped the pQTLs at four different P thresholds ($5x10^{-4}$, $5x10^{-5}$, $5x10^{-6}$, and $5x10^{-7}$) and extracted the P values of association with survival for each group (Fig. 7). We observed that there is a poor enrichment of pQTLs in variants associated with survival. All in all, our enrichment analysis of pQTLs for variants associated with susceptibility and survival showed that distinct genetic variation contributes to the three phenotypes studied here: inflammatory responses upon *Candida* stimulation, susceptibility to candidaemia and patient survival.

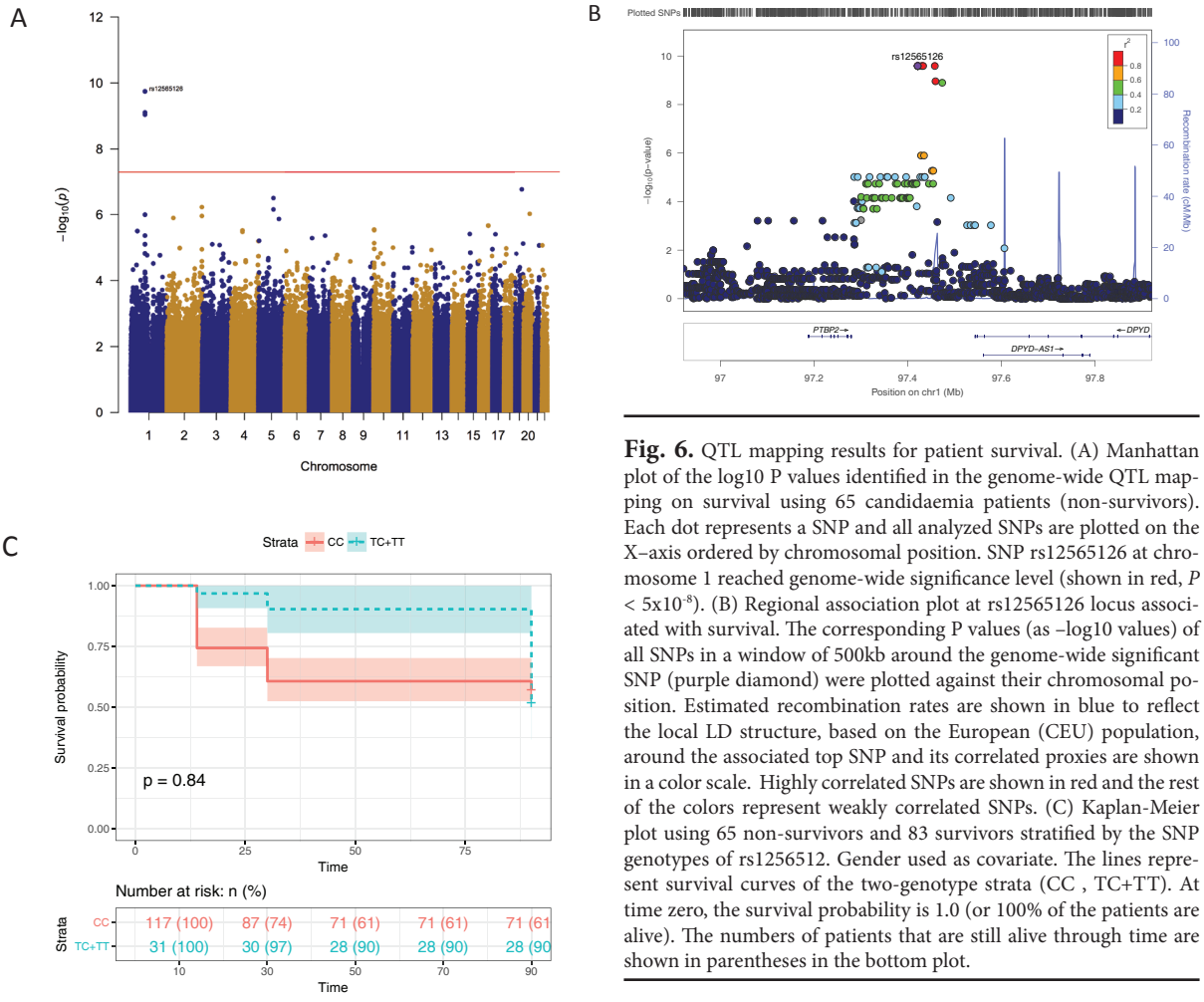



**Fig. 6.** QTL mapping results for patient survival. (A) Manhattan plot of the log10 P values identified in the genome-wide QTL mapping on survival using 65 candidaemia patients (non-survivors). Each dot represents a SNP and all analyzed SNPs are plotted on the X–axis ordered by chromosomal position. SNP rs12565126 at chromosome 1 reached genome-wide significance level (shown in red, $P < 5x10^{-8}$). (B) Regional association plot at rs12565126 locus associated with survival. The corresponding P values (as –log10 values) of all SNPs in a window of 500kb around the genome-wide significant SNP (purple diamond) were plotted against their chromosomal position. Estimated recombination rates are shown in blue to reflect the local LD structure, based on the European (CEU) population, around the associated top SNP and its correlated proxies are shown in a color scale. Highly correlated SNPs are shown in red and the rest of the colors represent weakly correlated SNPs. (C) Kaplan-Meier plot using 65 non-survivors and 83 survivors stratified by the SNP genotypes of rs1256512. Gender used as covariate. The lines represent survival curves of the two-genotype strata (CC , TC+TT). At time zero, the survival probability is 1.0 (or 100% of the patients are alive). The numbers of patients that are still alive through time are shown in parentheses in the bottom plot.

## Contribution of pQTLs that influence circulatory inflammatory proteins to complex infectious diseases

Given that the genome-wide pQTLs in response to *Candida* infection showed pleiotropy, we next investigated whether genetic variants that were previously associated with other infectious diseases are pQTLs. For this, we selected infectious diseases that showed at least ten independent associations per study in European populations, which are reported in the National Human Research Institute GWAS catalogue (https://www.ebi.ac.uk/gwas/). Next, we identified all pQTLs that were associated with inflammatory proteins at $P \leq 0.001$ and P value for heterogeneity $\geq 0.05$ and tested whether SNPs associated with infectious diseases or their prox-

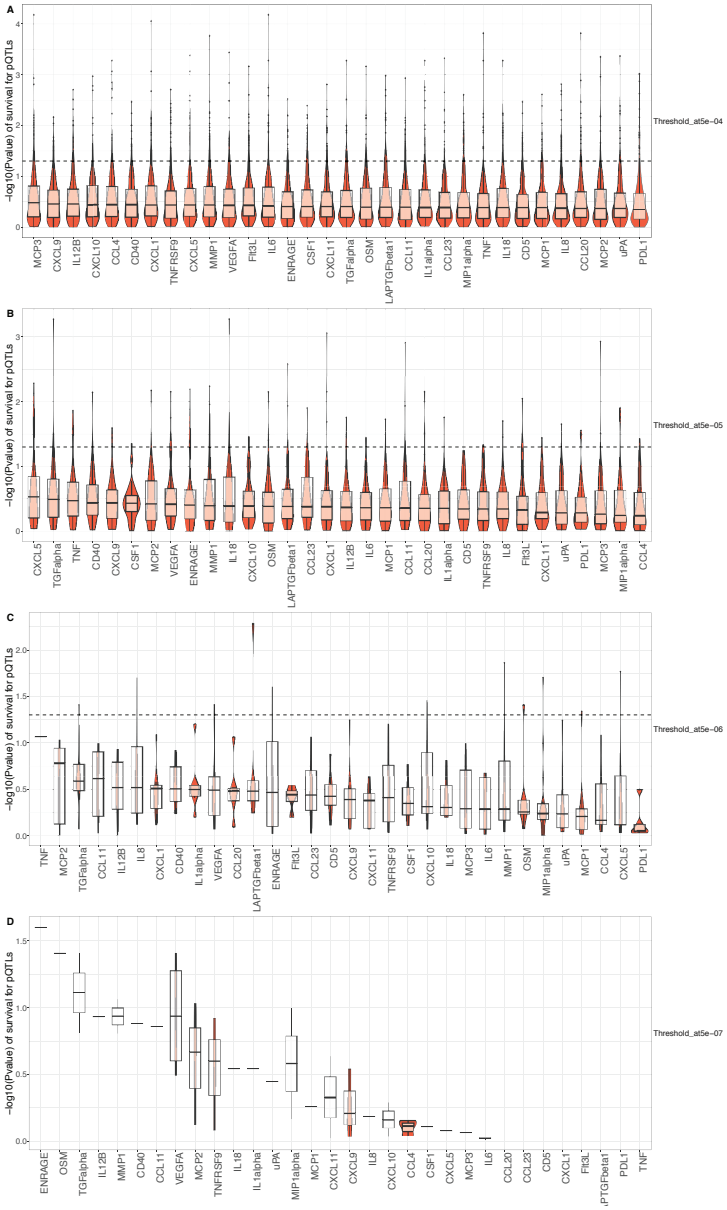



**Fig. 7.** A distinct genetic contribution between patient survival and inflammatory protein responses upon *Candida* stimulation was observed. Same approach was followed as described in Fig. 5. pQTLs at four different P thresholds ($5x10^{-4}$,$5x10^{-5}$, $5x10^{-6}$ and $5x10^{-7}$) showed a poor enrichment for variants associated with survival. Axis x show the 32 inflammatory proteins meaured in PBMCs in response to *Candida* stimulation and y axis represents the neg log10 of P values of association with survival. Violin plots show the distribution of P values of association with survival for each different threshold for pQTL effect. The black line within the box plots indicates the median and the dashed black line represents a P value of significance 0.05. pQTLs above the dashed line were considered to be significantly associated with patient survival.

ies (r2 > 0.8) were also pQTLs. We found seven intronic SNPs or proxies ($r^2 \geq 0.8$) associated with infectious diseases that were also pQTLs. Two loci were associated with cytomegalovirus antibody response, another two loci with severe influenza A (H1N1) infection, one locus with hepatitis B, one locus with acquired immunodeficiency syndrome (AIDS) progression and one locus associated with yeast infection (vulvovaginal candidiasis) (Table S9)[33,34]. When using more stringent

P values to call pQTLs, the number of disease-associated SNPs that were pQTLs was dropped to few SNPs (n < 10) as the number of available pQTLs was reduced. These preliminary results indicate that inflammatory proteins studied here have an important role as underlying mediators in other infectious diseases as well.

## Discussion

The incidence of opportunistic invasive fungal infections has been increased over the last decades. This increase can be explained by the use of aggressive and intensive chemotherapeutic regimens, of immunosuppressive therapy for autoimmune disorders, and of transplantation that have led to a rise in the number of susceptible human hosts[35]. The development of specific and mechanistically relevant biomarkers in *Candida* infections are important tools in identifying patients at high-risk and, thus, clinicians can provide the most beneficial prophylactic and/or treatment strategy. In this study, we performed a systematic proteomic analysis of 92 inflammatory-related proteins in blood circulation in response to *Candida* infection. To our knowledge, this is the first effort to date to evaluate a great number of inflammatory proteins and evaluate the role of host (non)-genetics in regulating inflammatory protein levels in *Candida* infection.

We observed significant differences in inflammatory responses in candidaemia patients compared to healthy individuals from Lifelines cohort. Also, we observed increased inter-individual variation in inflammatory responses in PBMCs stimulated with *C. albicans* compared to samples stimulated with RPMI medium. When comparing the clustering pattern of inflammatory responses between patients and healthy controls, we observed different clustering patterns, as expected. Of note, MMP-1 expression was stronger in patients compared to controls. MMP-1 belongs to the family of matrix metalloproteinases (MMPs) that act as processing enzymes that cleave most structural extracellular matrix (ECM) proteins but, also, cell surface receptors, cytokines, chemokines, clotting factors, cell-cell adhesion molecules, and other proteinases[36]. Previous studies indicated that MMPs play a critical role in the host defence against infection and, of particular interest, can be used as drug targets in infections caused by gram-negative bacteria and in septic shock[37]. Furthermore, we observed different clustering patterns of inflammatory responses between patients and *Candida*-stimulated supernatants of PBMCs, suggesting the presence of additional factors that contribute to the inflammatory responses in patients that are missing in PBMC fraction. These factors could be different cell types (e.g. neutrophils) or blood coagulation factors and platelets that may interact with the inflammatory proteins that are not present in the PBMC fraction.

To explain the observed inter-individual variations in inflammatory responses, we investigated the contribution of (non)-genetic factors to inflammatory responses. Although age and gender has a well-described influence on cytokine responses, correlation analysis of non-genetic factors measured in 500FG cohort with the inflammatory responses showed minimal effect, with few exceptions. Three inflammatory proteins (MMP-1, CXCL5 and CCL23) were significantly correlated with age ($P < 0.05$) and TNFRSF9, PD-L1, IL-1aplha, IL-18, IL-12B, EN-RAGE, CXCL9, and CSF-1 showed a significant (positive) correlation with gender ($P < 0.05$) (Figure S6 and Table S5).

In addition, we identified 10 independent significant pQTLs with MAF > 1% that reached genome-wide significance level ($P < 5x10^{-8}$, FDR 0.10) that influence CCL4, VEGF-A, IL-8, CXCL9, MCP-1, MCP-2 and MCP-3. These pQTLs are *trans*-pQTLs that influence inflammatory responses indirectly through regulatory loops. Pathway enrichment analysis of *Candida*-induced pQTLs with a *P* value < $9.99x10^{-4}$ showed an over-representation of genes involved in hemostasis and metabolism of lipids and lipoproteins (Figure S9), indicating the importance of hemostasis and lipid metabolism in anti-*Candida* inflammatory responses. Of note, we previously observed that stimulation of PBMCs with *C. albicans* yeast induces transcription of genes involved in inflammation, as expected, and immune-hemostasis interaction (*chapter 2*). Furthermore, we showed that *Candida*-induced cytokine-QTLs (cQTLs) in PBMCs that are associated with candidaemia susceptibility are enriched in lipid metabolism processes (*chapter 3*). Altogether, hemostasis and lipid metabolism seem to play a critical role in *Candida*-induced inflammatory responses and, thus, it would be interesting to further investigate their role in host immune defence against *C. albicans*.

To study the concordance of genetic effects on circulating proteins and expression levels in response to *C. albicans*, we also mapped *Candida*-response *cis*-eQTLs in PBMCs. The genome-wide significant pQTL rs392422, which influences IL-8, seems to affect the expression of a gene belonging to the non-coding RNA class, *CTC-281B15.1* (ENSG00000247572) (*P* 0.03). Also, three SNPs (rs191505409, rs187581857 and rs149399639), which are in high LD ≥ 0.8 with the genome-wide significant pQTL rs2501301, affect the expression of the non-coding RNA, *CDC42-IT1* (ENSG00000230068) in response to *C. albicans* (*P* 0.03, 0.009 and 0.009 respectively). The role of non-coding RNAs in regulating inflammation, including well-known mediators, such as TNFα, IL-1, IL-6, and IL-8 has been previously described[44], and, thus, we assume that regulation of inflammation during *C. albicans* infection can happen at the transcriptional level through non-coding RNAs and at translation level as well.

Another interesting finding of this study was that the genome-wide significant

*trans*-pQTLs showed pleiotropic effects. In particular, a locus at chromosome 3 and a second locus at chromosome 7 influence 17 and 11 distinct proteins respectively at $P < 1.56 \times 10^{-4}$ in the same direction. Both loci regulate different factors, such as various pro-inflammatory cytokines and chemokines, vascular endothelial growth factors (VEGF-A), pro-inflammatory mediators with effects on endothelial cells (OSM)[45] and transforming growth factors (TGF-α) (Table S8). This observation reflects the fact that an inflammatory response is the result of complex interactions between inflammatory and endothelial cells that can be regulated by the same genetic variation. In addition, we observed that SNPs associated with other complex infectious diseases, such as influenza A (H1N1) infection, hepatitis B, AIDS, vulvovaginal candidiasis and cytomegalovirus antibody response, were also pQTLs, highlighting the role of inflammatory proteins as critical mediators in many complex infectious diseases.

Moreover, we investigated the contribution of pQTLs to candidaemia susceptibility and patient survival. We observed a poor enrichment of pQTLs with SNPs associated with candidaemia susceptibility and survival. This finding suggests that distinct genetics determines circulatory inflammatory protein responses during *Candida* infection, candidaemia susceptibility and patient survival. This observation agrees with previous studies on other diseases, such as Crohn's disease, where distinct genetic contributions to prognosis and susceptibility were identified[46]. Another interesting finding of the within-case QTL mapping using 65 non-survivors revealed a genome-wide significant SNP (rs12565126) on chromosome 1 that is associated with survival. Despite the limited number of our survival cohort (n = 148), Kaplan-Meier analysis showed that individuals carrying the T allele (rs12565126) tend to survive longer compared to homozygotes for the C allele (Fig. 6C).

There are also limitations to our study. First, because of relatively small sample size, we cannot exclude that some of the pQTLs with rare allele frequencies were false positives in this analysis. In addition, the experimental setup of *ex vivo* PBMC stimulation for 24 hours with *C. albicans* yeast provides only the opportunity to study the interactions between immune cells, such as monocytes, T and B cells, missing the neutrophils and platelet fractions. Also, the time-dependent dynamic interactions are missing as PBMCs were stimulated for 24 hours. Lastly, it would be interesting to capture inflammatory responses upon stimulation with *C. albicans* hyphae as well, as the transition to hyphae contributes to *C. albicans* virulence. Despite these limitations, we provided an important groundwork by performing a comprehensive analysis of (non)-genetic variation that affects circulatory inflammatory responses in human blood circulation in response to *C. albicans* that can trigger critical future work to understand better the contribution of genetics to inflammatory responses against *C. albicans* and how it shapes sus-

ceptibility and survival in candidaemia patients.

## Materials and Methods

### Study populations

*Population-based cohorts*

*500FG*. To understand the variation in circulating inflammatory production (*in vitro*) in response to *C. albicans*, we used a population-based cohort, the 500FG, composed of 534 well-characterized healthy individuals of Western European ancestry. This cohort is part of the Human Functional Genomics Project (HFGP, see www.humanfunctionalgenomics.org). The 500FG cohort is composed of 237 males and 296 females all between 18 and 75 years old. The HFGP study was approved by the Ethical Committee of Radboud University Nijmegen, the Netherlands (no. 42561.091.12). Experiments were conducted according to the principles expressed in the Declaration of Helsinki and venous blood samples were drawn after written informed consent was obtained.

*Demographic Data collection in 500FG*. Volunteers, after donating blood in the hospital, received an extensive online questionnaire about lifestyle, diet, and disease history. Based on the results of this questionnaire, 45 volunteers were excluded for various reasons, e.g., they were under medication, non-European ancestry, or had a chronic disease. These exclusion criteria were taken to minimize false positive effects on the protein production capacity *in vitro*.

*Lifelines DEEP cohort*. Adult participants were enrolled from Lifelines, a large general population cohort in the Netherlands, after giving informed consent, following an institutional review board protocol approved by the University Medical Centre Groningen (Groningen, the Netherlands). The study design of Lifelines DEEP (i.e. DEEP: Detailed Extensive Examination of Participants) has been described elsewhere (http://lifelines.nl/researcher/biobank-lifelines/additional-studies/lifelines-deep).

*Patient cohort*. Adult candidaemia patients were enrolled after informed consent (or waiver as approved by the Institutional Review Board) at the Duke University Hospital (DUMC, Durham, North Carolina, USA). Enrollment occurred between January 2003 and January 2009 and, in total, 178 candidaemia cases and 175 case-matched controls of European ancestry were tested for disease association. The demographic and clinical characteristics of the candidaemia cohort have been described previously[47]. Patients must have had at least one positive blood culture for a *Candida* species, with the majority of them infected by *C. albicans*[47]. Case-

matched controls were recruited from the same hospital wards as candidaemia cases so that co-morbidities and clinical risk factors for infection were similar. One hundred forty-eight (148) candidaemia patients at DUMC were followed prospectively for up to 90 days and mortality dates were recorded for these individuals. Thirty-one out of 148 passed away within 14 days, 18 out of 148 passed away between 14 and 30 days and 16 out of 148 individuals passed away between 30 and 90 days (non-survivors). Eighty-three out of 148 individuals survived longer than the specified period of 90 days and we defined them as survivors in this study.

### Genotyping, quality control and imputation of the 500FG cohort

DNA obtained from the 500FG cohort was genotyped using the commercially available Illumina HumanOmniExpressExome-8 v1.0 SNP chip. Genotype calling was performed using Opticall 0.7.0 using default settings[48]. We applied quality control per sample to exclude samples with a call rate ≤ 0.99 and quality control per SNP to exclude variants with a Hardy-Weinberg equilibrium ≤ 0.0001, a call rate ≤ 0.99 and a minor allele frequency (MAF) ≤ 0.001. We identified 17 ethnic outliers by merging multi-dimensional scaling plots of samples with 1000 Genomes data, and these outliers were excluded from further analysis[12]. The quality control filters resulted in a dataset of 483 samples containing genotype information on 518,980 variants for further imputation. The strands and variants-identifiers were aligned to the reference Genome of the Netherlands (GoNL, Genome of the Netherlands Consortium, 2014[49]) dataset using Genotype Harmonizer[50] . The data were phased using SHAPEIT2 v2 using GoNL as a reference panel[51,52]. We selected SNPs that showed an INFO score ≥ 0.8 upon imputation for further cytokine QTL mapping.

### Genotyping, quality control and imputation of the candidaemia cohort

Isolated DNA obtained from case and control samples was genotyped using the commercially available SNP chips, HumanCoreExome-12 v1.0 and HumanCoreExome-24 v1.0 BeadChip from Illumina (https://www.illumina.com). Genotype calling was performed using Optical 0.7.0 using the default settings[48]. Strands of variants were aligned and identified against the 1000 Genome reference panel using Genotype Harmonizer[50]. We then imputed the samples on the Michigan imputation server using the human reference consortium as a reference panel, and we filtered variants with an $R^2$ of 0.3 for imputation quality[53]. After excluding imputed variants with a MAF < 0.10 and variants within the HLA region, we applied quality control per sample and removed 15 individuals (5 cases and 10 controls) due to excess/reduced heterozygosity and cryptic relatedness. Next, we applied SNP quality control filters to exclude SNPs with a MAF of < 0.05 and a Hardy-Weinberg equilibrium of $P < 1x10^{-6}$ in control samples only. We also

identified 25 ethnic outliers (12 cases and 13 controls) by multidimensional scaling analysis (MDS) (performed in PLINK on the N x N matrix of genome-wide IBS pairwise distance) of candidaemia patients and case-matched controls for the two first principle components (Figure S11A). The quantile-quantile (QQ) plot that shows the distribution of the observed –log10 for the whole-genome SNPs against the theoretical distribution of expected –log10 indicated no or little evidence of population stratification (Figure S11B). The genomic inflation factor based on median chisq was equal to 1 ($\lambda = 1$). Quality control filters resulted in a dataset of 161 cases and 152 disease-matched control samples containing genotype information on 5,326,313 SNP variants for further testing for association with candidaemia susceptibility.

**Genotyping, quality control and imputation of the Lifelines DEEP cohort**

Genotyping and genotype imputation of the Lifelines DEEP cohort have been published elsewhere[54]. DNA isolation was performed by the Qiagen robots using Autopure LS kits. Genotyping of DNA from the LifeLines DEEP cohort was performed using both the HumanCytoSNP-12 BeadChip and the ImmunoChip platforms (Illumina, San Diego, CA, USA). First, SNP quality control was applied independently for both platforms where SNPs were filtered on MAF above 0.001, a Hardy-Weinberg equivalent P value $>1e^{-4}$ and a call rate of >0.98 using Plink[55]. The genotypes generated from both platforms were merged into one dataset. After merging, SNPs were again filtered on MAF 0.05 and a call rate of 0.98, resulting in a total of 379,885 genotyped SNPs. Next, genotypes were imputed using the Genome of the Netherlands (GoNL) reference panel[56–58]. The merged genotypes were pre-phased using SHAPEIT2 and aligned to the GoNL reference panel using Genotype Harmonizer (http://www.molgenis.org/systemsgenetics/) in order to resolve strand issues[50]. Imputation was performed using IMPUTE2 version 2.3.0 against the GoNL reference panel[59].

**PBMC collection and *Candida* stimulation experiments**

Venous blood from the cubital vein of healthy volunteers was drawn in 10 ml EDTA Monoject tubes (Medtronic, Dublin) after obtaining written informed consent. PBMC isolation was performed as previously described[60]. In short, the PBMC fraction was obtained using density centrifugation of EDTA blood diluted 1:1 in pyrogen-free saline over Ficoll-Paque (Pharmacia Biotech, Uppsala). Cells were then washed twice in saline and suspended in RPMI medium supplemented with gentamicin (10 mg/mL), L-glutamine (10 mM) and pyruvate (10 mM). Addition of antibiotics such as gentamycin is a standard method used to avoid contamination of cultures, and it does not influence the ability to induce cytokine production by PBMCs or macrophages (data not shown). PBMCs were counted in a Coulter counter (Beckman Coulter, Pasadena) and re-suspended in a con-

centration of $5x10^6$ cells/mL. A total of $5x10^5$ PBMCs were added in 100 µL to round-bottom 96-well plates (Greiner) and incubated with 100 µL of stimulus (heat killed *Candida* albicans yeast of strain ATCC MYA-3573, UC 820, $1x10^6$/mL) or RPMI medium. After 24 hours, the supernatants were collected and stored at −20°C until assayed using commercial available ELISA kits (PeliKine Compact or R&D Systems) or, OLINK technology for high-throughput protein profiling (https://www.olink.com/).

### Proteomic profiling of circulating inflammatory proteins

Supernatant samples of PBMCs stimulated with *C. albicans* yeast (obtained from 500FG and Lifelines DEEP cohorts) and plasma samples obtained from candidaemia patients and healthy individuals from Lifelines cohort were analyzed using the proximity extension assay (PEA). We selected the Olink® inflammatory panel (Olink Bioscience AB, Uppsala, Sweden). This panel includes 92 inflammatory proteins. The protein analysis is reported as Normalized Protein Expression (NPX) values, which are Cq values normalized by the subtraction of values for extension control, as well as inter-plate control. The scale is shifted using a correction factor (normal background noise) and reported in Log2 scale. The normality test performed on both raw and log-transformed data using Shapiro-Wilk normality test, and a $P > 0.05$ was used as a threshold for normal distribution. We further validated the correlation between protein concentrations of IL-6 measured by OLINK technology in log2 picogram/ml (NPX values) and by ELISA (Figure S12). We used Spearman's correlation as the measure of similarity between the two meausurements.

### Clustering Analysis

To analyze the similarities and dissimilarities between the measured inflammatory proteins in PBMC supernatants and plasma samples, we performed unsupervised hierarchical clustering using Spearman's correlation as the measure of similarity. To identify whether different cell counts in PBMC fraction have an impact on the circulating proteins, we obtained cell count data measured by FACS for total lymphocytes, T cells, B cells, monocytes and NK-cells from 487 individuals from the 500FG cohort, as previously described[17]. Then, we tested the correlation between cell counts in log2 scale with NPX using Spearman's correlation as the measure of similarity. All statistical analyses regarding the protein data were performed in R version 3.3.2 (https://www.R-project.org/).

### Protein QTL mapping

We performed a joint analysis of pQTLs by combining two different population-based cohorts (500FG and Lifelines DEEP cohorts). Lack of protein measurements for all available individuals restricted us to select 360 individuals from

500FG and 76 samples from Lifelines DEEP cohort for whom both genotype and protein data were available. The number of proteins measured in at least 90% of the PBMC samples in 500FG (Table S3) and Lifelines (Table S7) was 32 and 46 respectively. For mapping, we selected proteins that were detected in both cohorts in at least 90% of the samples and, upon quality control of protein distribution (Figure S7 and S8), we obtained a total of 32 inflammatory proteins (Table S3 and S7). Shapiro test was used to test for normality under the null hypothesis that the protein distribution is normal. If P value < 0.05, the distribution is non-normal. The majority of proteins expressed in PBMCs from 500FG followed a non-Gaussian distribution, with the exception of three proteins, CD40, VEGF-A and TNFRSF9 (Figure S7 and Table S3). Non-Gaussian distribution was also observed for proteins expressed in PBMCs from Lifelines DEEP cohort, with the exception of few proteins (Figure S8 and Table S7).

To correct the protein distributions for QTL mapping, we used a linear model adjusting for age and gender. NPX values of protein levels were used and the correlation between protein production and genotype was tested independently for both cohorts using the R-package Matrix-eQTL[61]. To jointly test the effect of genetic variation on protein levels measured in both 500FG and Lifelines DEEP cohorts, we used METAL (http://www.sph.umich.edu/csg/abecasis/metal/) and removed pQTLs with opposite direction and those that show significant heterogeneity ($P < 0.05$) between the two cohorts[62]. We considered $P < 5x10^{-8}$ to be the threshold for genome-wide significant pQTLs.

## Measurements of circulating mediators & correlation analysis of non-genetic factors

The circulating mediators resistin, leptin, adiponectin, CRP and alpha-1 antitrypsin (AAT) were measured in EDTA plasma using the R&D Systems DuoSet ELISA kits following the Manufacturer's protocol. To identify the effect of non-genetic factors, such as age, gender, BMI, oral contraceptive use and circulating mediators (resistin, adiponectin, alpha-1 Antitrypsin, leptin and CRP) on levels of circulatory inflammatory proteins in 500FG, we performed a rank-based regression analysis using "Rfit", as previously described[16].

## Pleiotropy

To test all reported genome-wide significant pQTLs for pleiotropy, we initially inspected all genome-wide significant *trans*-pQTLs whether they show association with multiple distinct proteins. To correct for the multiple testing burden for all genome-wide pQTLs, we next inspected all genome-wide significant *trans*-pQTLs that had at least two associations with distinct proteins at $P < 0.05/(10*32)$

$= 1.56x10^{-4}$. This cut-off represents a conservative approach to the multiple testing burden for all identified GWAS SNPs (n = 10) with all protein traits (n = 32). The resulting association matrix was then clustered and visualized based on the negative log10 of the P values of association. For clustering analysis, we used an unsupervised clustering approach using Spearman's correlation as the measure of similarity.

### RNA sequencing, expression analysis and eQTL mapping

RNA sequencing data from PBMCs of eight individuals stimulated by *C. albicans* were generated and published elsewhere[63]. Sequencing reads were mapped to the human genome using STAR (version 2.3.0)[64]. The aligner was provided with a file containing junctions from Ensembl GRCh37.71. Htseq-count of the Python package HTSeq (version 0.5.4p3) was used (The HTSeq package, http:// www-huber.embl.de/users/anders/HTSeq/doc/overview.html) to quantify the read counts per gene based on annotation version GRCh37.71, using the default union-counting mode. Differentially expressed genes were identified by statistics analysis using DESeq2 package from bioconductor[65]. The statistically significant threshold (FDR $P \leq 0.05$ and fold change $\geq 2$) was applied. To assess the effect of genetic variation on gene expression, gender and age were included as known covariates in a linear model. All eQTL mapping was done using Matrix-eQTL[61].

### Genome-wide case-control association analysis with susceptibility to candidaemia

The associations between the genetic variants and candidaemia susceptibility were described in *chapter 3*. Briefly, associations were tested by Fisher's exact test using PLINK 1.9 (www.cog-genomics.org/plink/1.9/)[66]. The genomic inflation factor ($\lambda = 1$) indicated that there was no population stratification effect (Figure S11B). A P value significance threshold of $< 5x10^{-8}$ was set to call genome-wide significant associations.

### Within-case QTL mapping and survival analysis

To determine genetic variants that are associated with survival, we performed a within-case QTL mapping on a genome-wide scale using 65 candidaemia patients for which mortality dates were recorced (non-survivors). In detail, 31 patients passed away within 14 days, 18 passed away between 14 and 30 days and 16 passed away between 30 and 90 days. Multidimensional scaling analysis (MDS) of candidaemia cohort, including non-survivors, showed genetic homogeneity between non-survivors (Figure S11A). For QTL mapping, we assigned the actual day of survival as quantitative trait using PLINKv1.09[55]. Gender was included as known covariate in a linear model to map QTLs that influence patient survival. Regional

association plots of loci associated with survival were prepared using a web-based plotting tool, LocusZoom[67]. Lastly, survival analysis was done using a cohort of 148 candidaemia patients, of which 65 were non-survivors and 83 survivors . We stratified the 65 non-survivors and 83 survivors based on the SNP genotype and evaluated differences between SNP genotype strata by using a log-rank test of the R package survival. Gender was included as covariate in the survival analysis. Visualization was done with the R package survminer.

### Within-case genome-wide association testing for patient survival

To identify genetic variants that are associated with 14-, 30-, and 90-day survival, we run a within-case genome-wide association analysis using a survival cohort of 148 candidaemia patients. To test for association (1) with 14-day survival, we used 31 non-survivors that passed away within 14 days and 83 survivors, with (2) 30-day survival, we used 49 non-survivors that passed away within 30 days and 83 survivors, and with (3) 90-day survival, we used 65 non-survivors that passed away within 90 days and 83 survivors. Gender and the first two principal components calculated in MDS analysis (Figure S11A) were included as covariates in an additive model using SNPTEST v.2.5.4-beta3[52,68,69].

### Enrichment analysis of pQTLs for genetic variants associated with candidaemia susceptibility and patient survival

We first called pQTLs at four different thresholds ($5x10^{-4}$, $5x10^{-5}$, $5x10^{-6}$, and $5x10^{-7}$). Next, we intersected the pQTLs with variants associated with candidaemia susceptibility, and those associated with survival (identified by quantitative analysis). We then extracted P values of association with susceptibility and survival for the pQTLs binned at the four different thresholds. All graphs were visualized using the R package ggplot2.

### Extraction of SNPs associated with infectious diseases

SNPs associated with a number of infectious diseases that showed $P < 1x10^{-6}$ were extracted using the GWAS catalogue (https://www.genome.gov/gwastudies/). We selected diseases that were studied in European populations, and for which at least ten independent associations were reported per study. We then intersected the SNPs associated with infectious diseases and their proxies ($r^2 \geq 0.8$) with pQTLs that showed $P < 0.001$ in this study.

### Acknowledgments

The authors thank all volunteers from the 500 Functional Genomics cohort (500FG) of the Human Functional Genomics Project and Lifelines (DEEP) cohort for participation in the study. The authors would like to thank K. McIntyre

for editing the final text. This work was supported by a Research Grant [2017] of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) and Hypatia tenure track grant to V.K., European Research Council (ERC) Consolidator Grant [FP/2007-2013/ERC grant 2012-310372] and a Netherlands Organization for Scientific Research (NWO) Spinoza prize grant [NWO SPI 94-212] to M.G.N., an ERC Advanced grant [FP/2007-2013/ERC grant 2012-322698] and an NWO Spinoza prize grant [NWO SPI 92-266] to C.W., a European Union Seventh Framework Programme grant (EU FP7) TANDEM project [HEALTH-F3-2012-305279] to C.W. and V.K., Y.L. and M.O. were supported by a VENI grant (863.13.011 and 016.176.006) from the Netherlands Organization for Scientific Research (NWO). V.M. is supported by a PhD scholarship from Graduate School of Medical Sciences, University of Groningen, the Netherlands.

## Declaration of Interests

The authors declare no competing interests.

# References

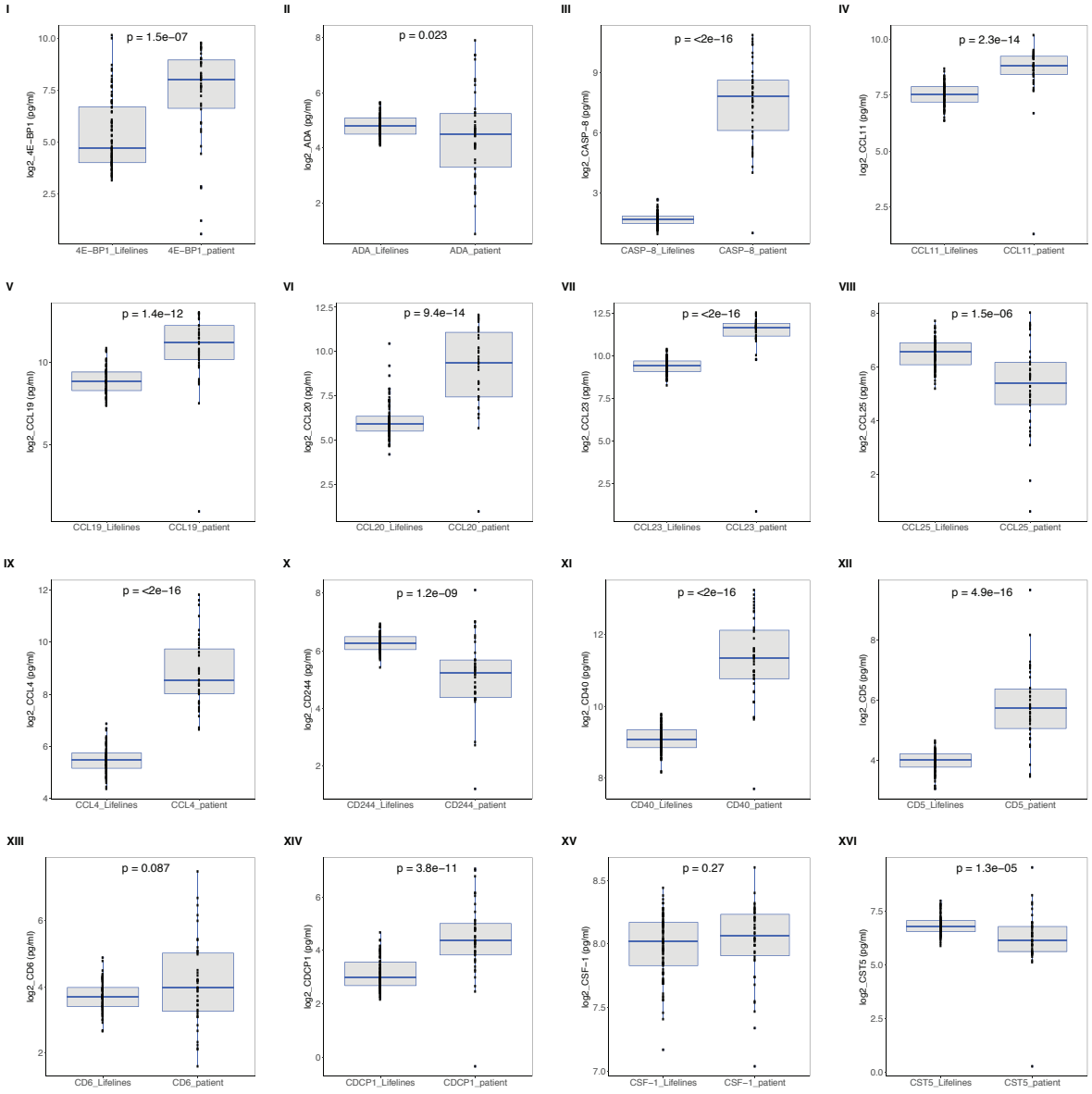
1. Centers for Disease Control and Prevention. Monitoring hospital-acquired infections to promote patient safety - United states, 1990-1999. Conn. Med. 64, 213–216 (2000).
2. Edmond, M. B. et al. Nosocomial bloodstream infections in United States hospitals: a three-year analysis. Clin. Infect. Dis. 29, 239–44 (1999).
3. Campion, E. W., Kullberg, B. J. & Arendrup, M. C. Invasive Candidiasis. N. Engl. J. Med. 373, 1445–1456 (2015).
4. Richardson, J. P. & Moyes, D. L. Adaptive immune responses to Candida albicans infection. Virulence 6, 327–337 (2015).
5. Netea, M. G., Joosten, L. A., van der Meer, J. W., Kullberg, B. J. & van de Veerdonk, F. L. Immune defence against Candida fungal infections. Nat Rev Immunol 15, 630–642 (2015).
6. Sironi, M. et al. Benzydamine inhibits the release of tumor necrosis factor-alpha and monocyte chemotactic protein-1 by Candida albicans-stimulated human peripheral blood cells. Int.J.Clin.Lab Res. 27, 118–122 (1997).
7. Hachicha, M., Rathanaswami, P., Naccache, P. H. & McColl, S. R. Regulation of chemokine gene expression in human peripheral blood neutrophils phagocytosing microbial pathogens. J Immunol 160, 449–454 (1998).
8. Huang, C., Levitz, S. M. Stimulation of macrophage inflammatory protein-1α, macrophage inflammatory protein-1β, and RANTES by Candida albicans and Cryptococcus neoformans in peripheral blood mononuclear cells from persons with and without human immunodeficiency virus infection. J. Infect. Dis. 181, 791–794 (2000).
9. Slaats J, Ten Oever J, van de Veerdonk FL, N. M. IL-1β/IL-6/CRP and IL-18/ferritin: Distinct Inflammatory Programs in Infections. PLoS Pathog 12, e1005973 (2016).
10. Mandel, S. A., Morelli, M., Halperin, I. & Korczyn, A. D. Biomarkers for prediction and targeted prevention of Alzheimer's and Parkinson's diseases: evaluation of drug clinical efficacy. EPMA J. 1, 273–292 (2010).
11. Walzl, G. et al. Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. Lancet Infect. Dis. 18, e199–e210 (2018).
12. Li, Y. et al. A functional genomics approach to understand variation in cytokine production in humans. Cell 167, 1099–1110.e14 (2016).
13. Li, Y. et al. Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. Nat. Med. (2016). doi:10.1038/nm.4139
14. Matzaraki, V. et al. An integrative genomics approach identifies novel pathways that influence candidaemia susceptibility. PLoS One 12, (2017).
15. Flevari A., Theodorakopoulou M., Velegraki A., Armaganidis A., D. G. Treatment of invasive candidiasis in the elderly: a review. Clin Interv Aging 8, 1199–1208 (2013).
16. ter Horst, R. et al. Host and environmental factors influencing individual human cytokine responses. Cell 167, 1111–1124.e13 (2016).
17. Aguirre-Gamboa, R. et al. Differential effects of environmental and genetic factors on T and B cell immune traits. Cell Rep. 17, 2474–2487 (2016).
18. Giefing-Kröll, C., Berger, P., Lepperdinger, G. & Grubeck-Loebenstein, B. How sex and age affect immune responses, susceptibility to infections, and response to vaccination. Aging Cell 14, 309–321 (2015).
19. Mangino, M., Roederer, M., Beddall, M. H., Nestle, F. O. & Spector, T. D. Innate and adaptive immune traits are differentially affected by genetic and environmental factors. Nat. Commun. 8, 13850 (2017).
20. Brodin, P. et al. Variation in the human immune system is largely driven by non-heritable influences. Cell 160, 37–47 (2015).
21. Silva-García O., Valdez-Alarcón J.J., B.-A. V. M. The Wnt/β-catenin signaling pathway controls the inflammatory response in infections caused by pathogenic bacteria. Mediators Inflamm. 2014, 7 (2014).
22. Brandenburg, J. & Reiling, N. The Wnt blows: On the functional role of Wnt signaling in mycobacterium tuberculosis infection and beyond. Frontiers in Immunology 7, 635 (2016).
23. Lamkanfi, M. & Kanneganti, T. D. Caspase-7: A protease involved in apoptosis and inflammation. International Journal of Biochemistry and Cell Biology 42, 21–24 (2010).
24. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue

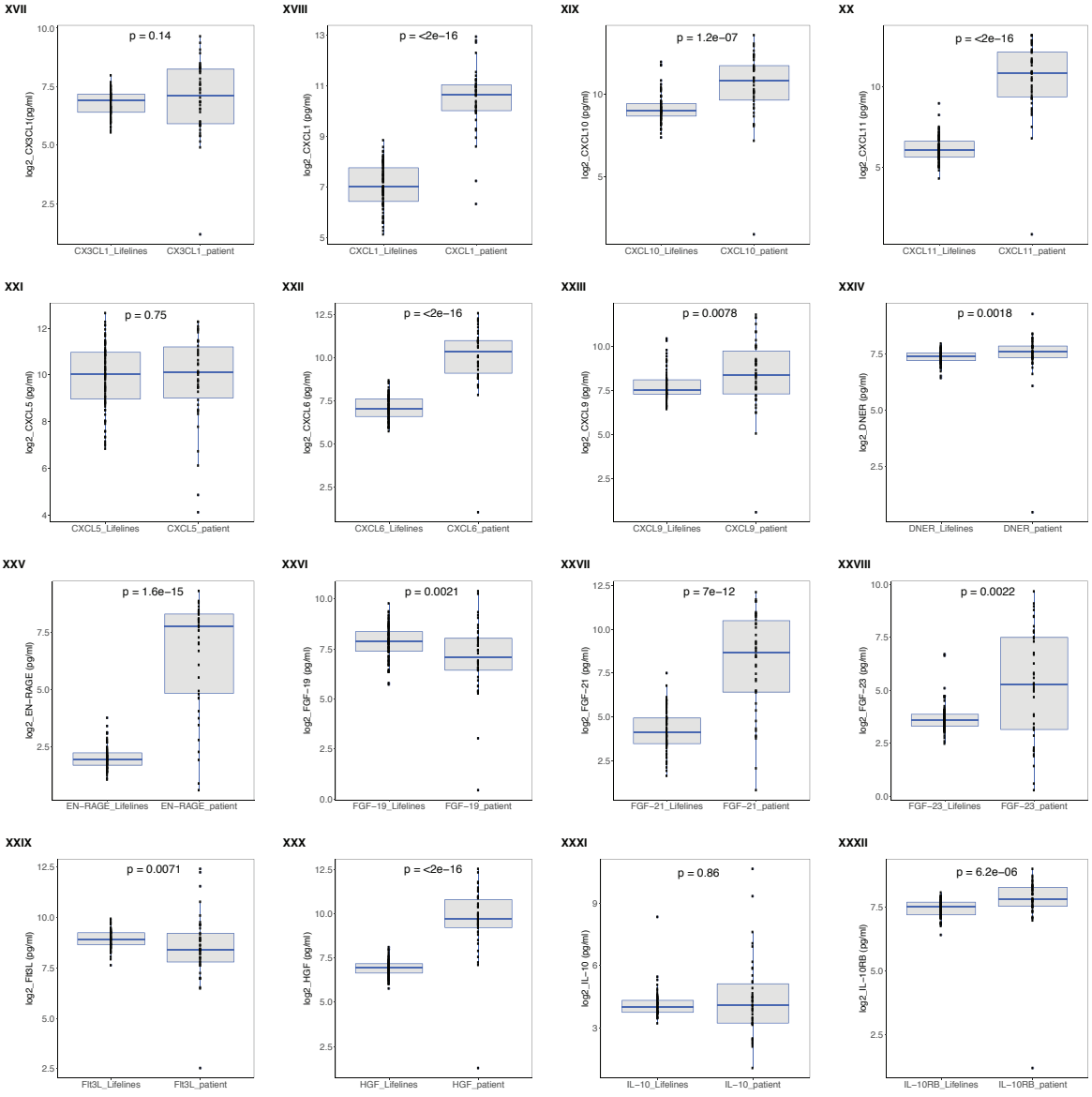
gene regulation in humans. Science 348, 648–660 (2015).
25. Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–43 (2013).
26. Minton, K. Natural killer cells: A TACTILE restraint. Nature Reviews Immunology 14, 285 (2014).
27. Lord, C. C., Thomas, G. & Brown, J. M. Mammalian alpha beta hydrolase domain (ABHD) proteins: Lipid metabolizing enzymes at the interface of cell signaling and energy metabolism. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids 1831, 792–802 (2013).
28. Westra, H. J. & Franke, L. From genome to function by studying eQTLs. Biochimica et Biophysica Acta - Molecular Basis of Disease 1842, 1896–1902 (2014).
29. Kuparinen, T. et al. Genome-wide association study does not reveal major genetic determinants for anti-cytomegalovirus antibody response. Genes Immun. 13, 184–190 (2012).
30. Tian, C. et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat. Commun. 8, (2017).
31. Troyer, J. L. et al. Genome-wide association study implicates PARD3B-based AIDS restriction. J. Infect. Dis. 203, 1491–1502 (2011).
32. Garcia-Etxebarria, K. et al. No major host genetic risk factor contributed to A(H1N1)2009 influenza severity. PLoS One 10, e0141661 (2015).
33. Liu J., Xia H., Kim M., Xu L., Li Y., Zhang L., Cai Y., Norberg H.V., Zhang T., Furuya T., Jin M., Zhu Z., Wang H., Yu J., Li Y., Hao Y., Choi A., Ke H., Ma D., Y. J. Beclin1 controls the levels of p53 by regulating the deubiquitination activity of USP10 and USP13. Cell 147, 223–34 (2011).
34. Liu, Y. et al. USP13 antagonizes gp78 to maintain functionality of a chaperone in ER-associated degradation. Elife 2014, e01369 (2014).
35. Singh, N. Trends in the epidemiology of opportunistic fungal infections: predisposing factors and the impact of antimicrobial use practices. Clin. Infect. Dis. 33, 1692–1696 (2001).
36. Stamenkovic, I. Extracellular matrix remodelling: the role of matrix metalloproteinases. J. Pathol. 200, 448–464 (2003).
37. Vanlaere, I. & Libert, C. Matrix metalloproteinases as drug targets in infections caused by gram-negative bacteria and in septic shock. Clinical Microbiology Reviews 22, 224–239 (2009).
38. Marques-Rocha, J. L. et al. Noncoding RNAs, cytokines, and inflammation-related diseases. FASEB J. 29, 3595–3611 (2015).
39. Modur, V. et al. Oncostatin M is a proinflammatory mediator: In vivo effects correlate with endothelial cell expression of inflammatory cytokines and adhesion molecules. J. Clin. Invest. 100, 158–168 (1997).
40. Lee, J. C. et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. Nat. Genet. 49, 262–268 (2017).
41. Kumar, V. et al. Immunochip SNP array identifies novel genetic variants conferring susceptibility to candidaemia. Nat. Commun. 5, 4675 (2014).
42. Shah, T. S. et al. OptiCall: A robust genotype-calling algorithm for rare, low-frequency and common variants. Bioinformatics 28, 1598–1603 (2012).
43. Collection, S. & Genome, T. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. 46, 1–95 (2014).
44. Deelen, P. et al. Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. BMC Res. Notes 7, 901 (2014).
45. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods 10, 5–6 (2013).
46. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. G3:Genes|Genomes|Genetics 1, 457–470 (2011).
47. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48, 1279–1283 (2016).
48. Ricaño-Ponce, I. et al. Refined mapping of autoimmune disease associated genetic variants with gene expression suggests an important role for non-coding RNAs. J. Autoimmun. 68, 62–74 (2016).
49. Purcell, S. et al. PLINK: A Tool Set for Whole-Genome Association and Popula-

tion-Based Linkage Analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
50. Boomsma, D. I. et al. The Genome of the Netherlands: Design, and project goals. Eur. J. Hum. Genet. 22, 221–227 (2014).
51. Deelen, P. et al. Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of the Netherlands'. Eur. J. Hum. Genet. 22, 1321–1326 (2014).
52. Consortium, T. G. of the N. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. 46, 1–95 (2014).
53. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, (2009).
54. Oosting, M. et al. Borrelia-induced cytokine production is mediated by spleen tyrosine kinase (Syk) but is Dectin-1 and Dectin-2 independent. Cytokine 76, 465–472 (2015).
55. Shabalin, A. A. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353–1358 (2012).
56. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
57. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, (2014).
59. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
60. Pruim, R. J. et al. LocusZoom: Regional visualization of genome-wide association scan results. in Bioinformatics 27, 2336–2337 (2011).
61. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007).
62. Wellcome, T., Case, T. & Consortium, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–78 (2007).

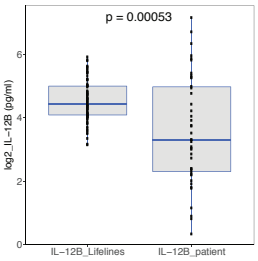
# Supplementary Information

**Figure S1:** Protein expression of inflammatory proteins in plasma of candidaemia patients compared to plasma from healthy individuals (Lifelines cohort). P values were calculated using Wilcoxon rank sum test, and $P < 0.05$ was considered to be the threshold for significance.

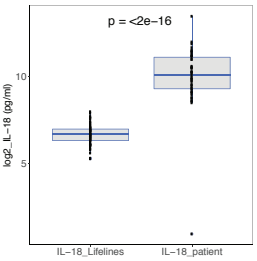



XVII
p = 0.14
log2_CX3CL1(pg/ml)
CX3CL1_Lifelines
CX3CL1_patient
XVIII
p = <2e−16
log2_CXCL1 (pg/ml)
CXCL1_Lifelines
CXCL1_patient
XIX
p = 1.2e−07
log2_CXCL10 (pg/ml)
CXCL10_Lifelines
CXCL10_patient
XX
p = <2e−16
log2_CXCL11 (pg/ml)
CXCL11_Lifelines
CXCL11_patient
XXI
p = 0.75
log2_CXCL5 (pg/ml)
CXCL5_Lifelines
CXCL5_patient
XXII
p = <2e−16
log2_CXCL6 (pg/ml)
CXCL6_Lifelines
CXCL6_patient
XXIII
p = 0.0078
log2_CXCL9 (pg/ml)
CXCL9_Lifelines
CXCL9_patient
XXIV
p = 0.0018
log2_DNER (pg/ml)
DNER_Lifelines
DNER_patient
XXV
p = 1.6e−15
log2_EN−RAGE (pg/ml)
EN−RAGE_Lifelines
EN−RAGE_patient
XXVI
p = 0.0021
log2_FGF−19 (pg/ml)
FGF−19_Lifelines
FGF−19_patient
XXVII
p = 7e−12
log2_FGF−21 (pg/ml)
FGF−21_Lifelines
FGF−21_patient
XXVIII
p = 0.0022
log2_FGF−23 (pg/ml)
FGF−23_Lifelines
FGF−23_patient
XXIX
p = 0.0071
log2_Flt3L (pg/ml)
Flt3L_Lifelines
Flt3L_patient
XXX
p = <2e−16
log2_HGF (pg/ml)
HGF_Lifelines
HGF_patient
XXXI
p = 0.86
log2_IL−10 (pg/ml)
IL−10_Lifelines
IL−10_patient
XXXII
p = 6.2e−06
log2_IL−10RB (pg/ml)
IL−10RB_Lifelines
IL−10RB_patient

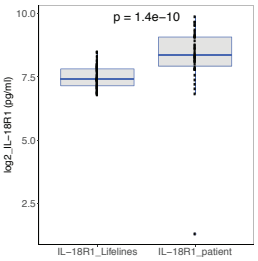
XXXIII



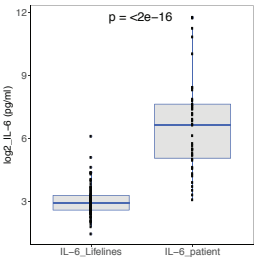

XXXIV



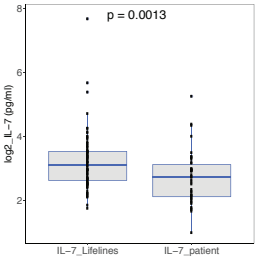

XXXV



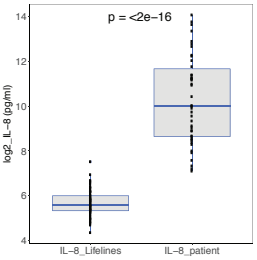

XXXVI



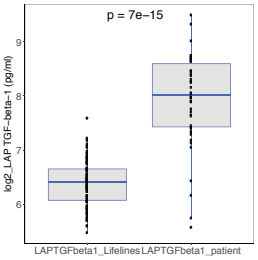

XXXVII



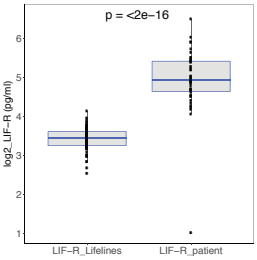

XXXVIII



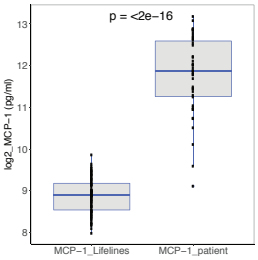

XXXIX



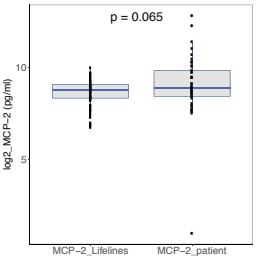

XL



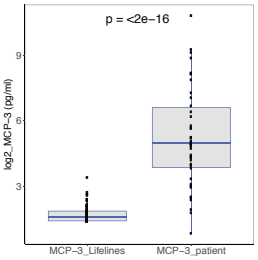

XLI



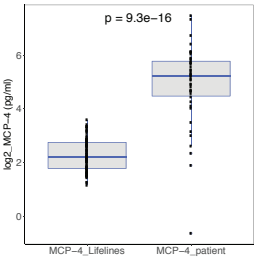

XLII



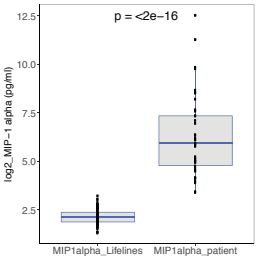

XLIII



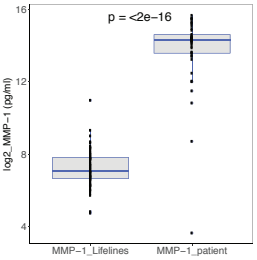

XLIV



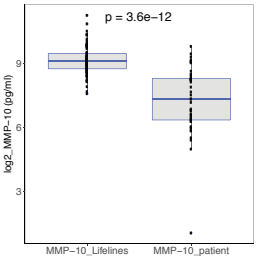

XLV



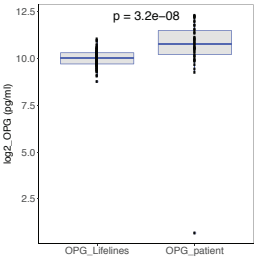

XLVI



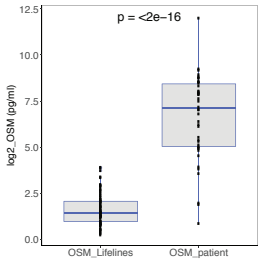

XLVII



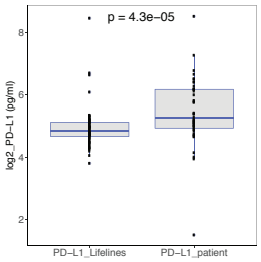

XLVIII



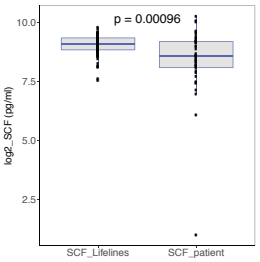
XLIX



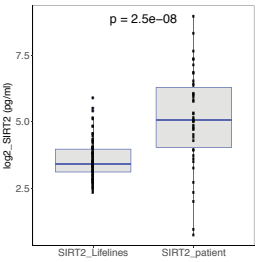

L



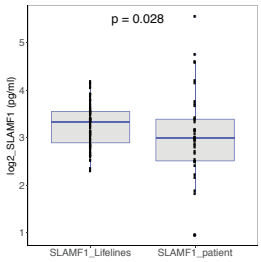

LI



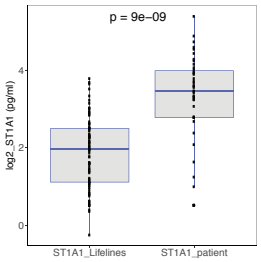

LII



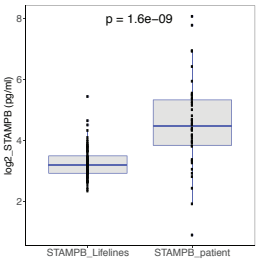

LIII



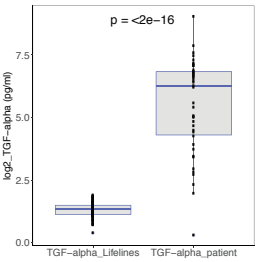

LIV



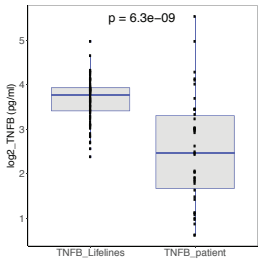

LV



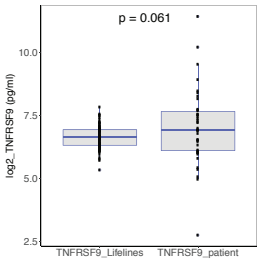

LVI



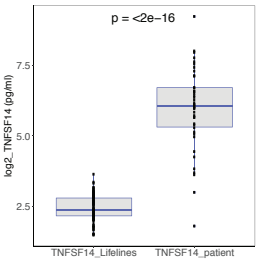

LVII



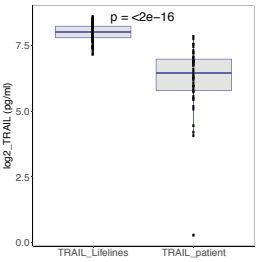

LVIII



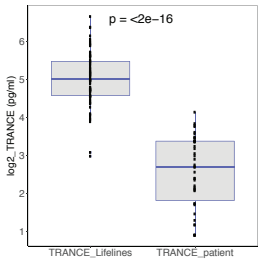

LIX



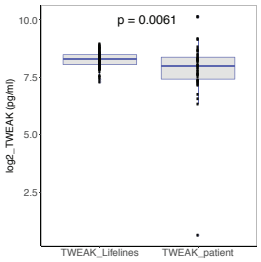

LX



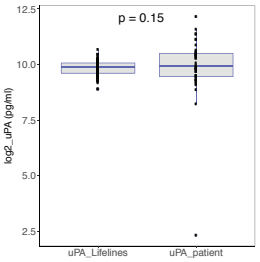

LXI



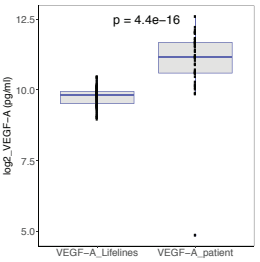

LXII




LXIII




LXIV



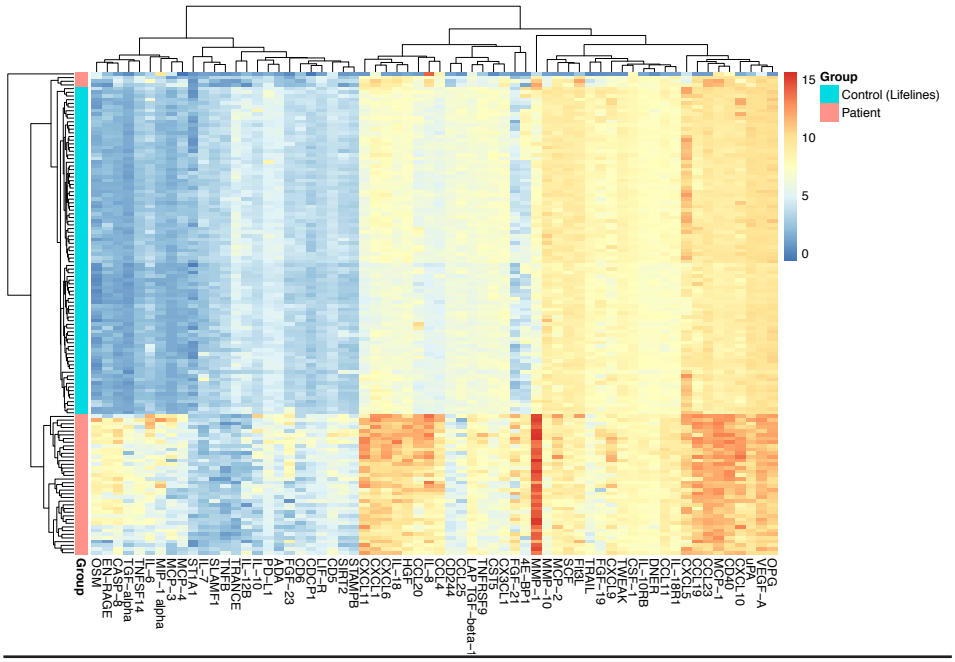
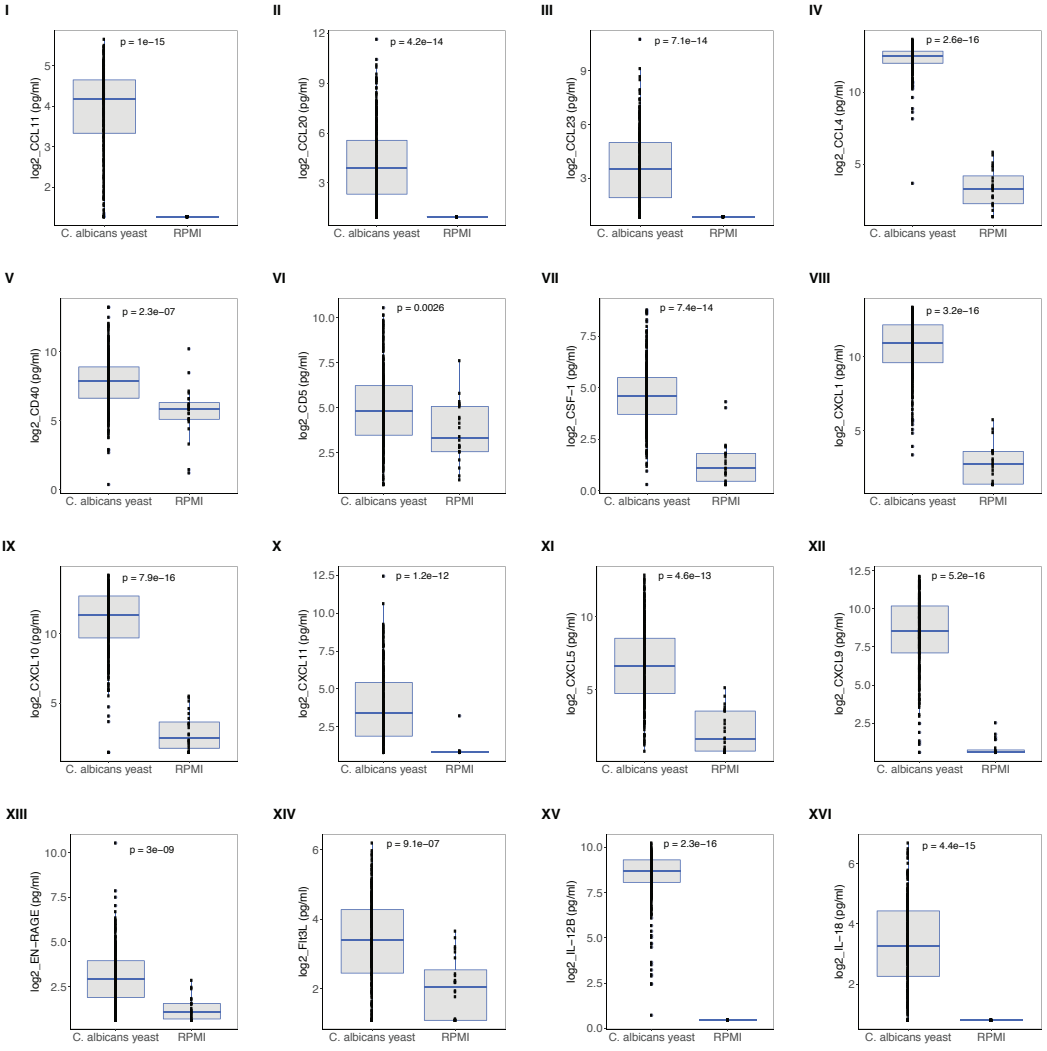
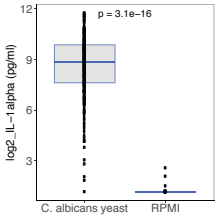



**Figure S2:** Unsupervised clustering of inflammatory responses in patients and healthy controls from Lifelines cohort revealed distinct clustering patterns among plasma proteins between patient and control samples. Unsupervised hierarchical clustering was performed using the measurements of 64 proteins that were expressed in both plasma of patients and healthy controls. These proteins were measured in at least 90% of the plasma samples in candidaemia patients and healthy individuals.

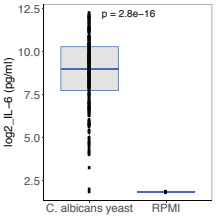
**Figure S3:** Protein expression of inflammatory proteins that were detected in at least 90% of the supernatants of PBMC samples stimulated with *C. albicans* yeast compared to samples stimulated with RPMI medium. P values were calculated using Wilcoxon rank sum test, and $P < 0.05$ was considered to be the threshold for significance. Samples were stimulated with *C. albicans* yeast or RPMI medium for 24 hours, as described in Material and Methods section.



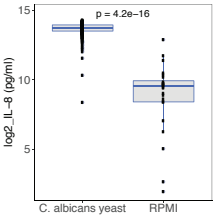
XVII



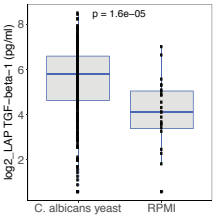

XVIII



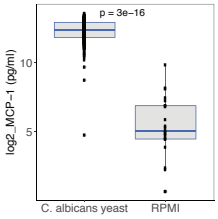

XIX



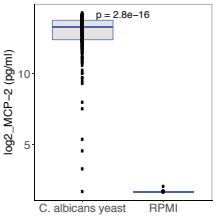

XX



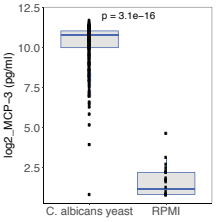

XXI



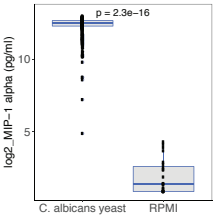

XXII



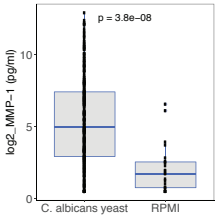

XXIII



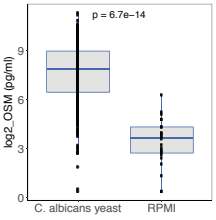

XXIV



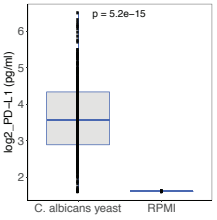

XXV



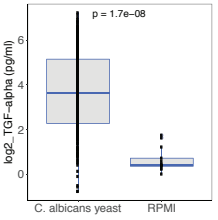

XXVI



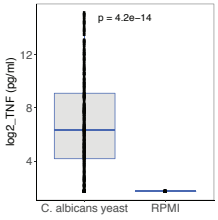

XXVII



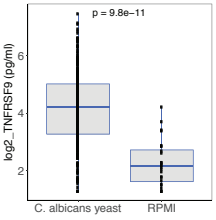

XXVIII



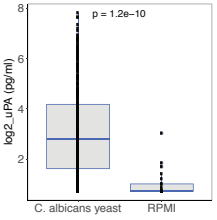

XXIX



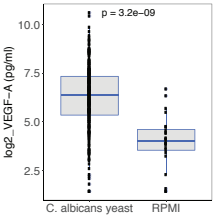

XXX




XXXI




XXXII



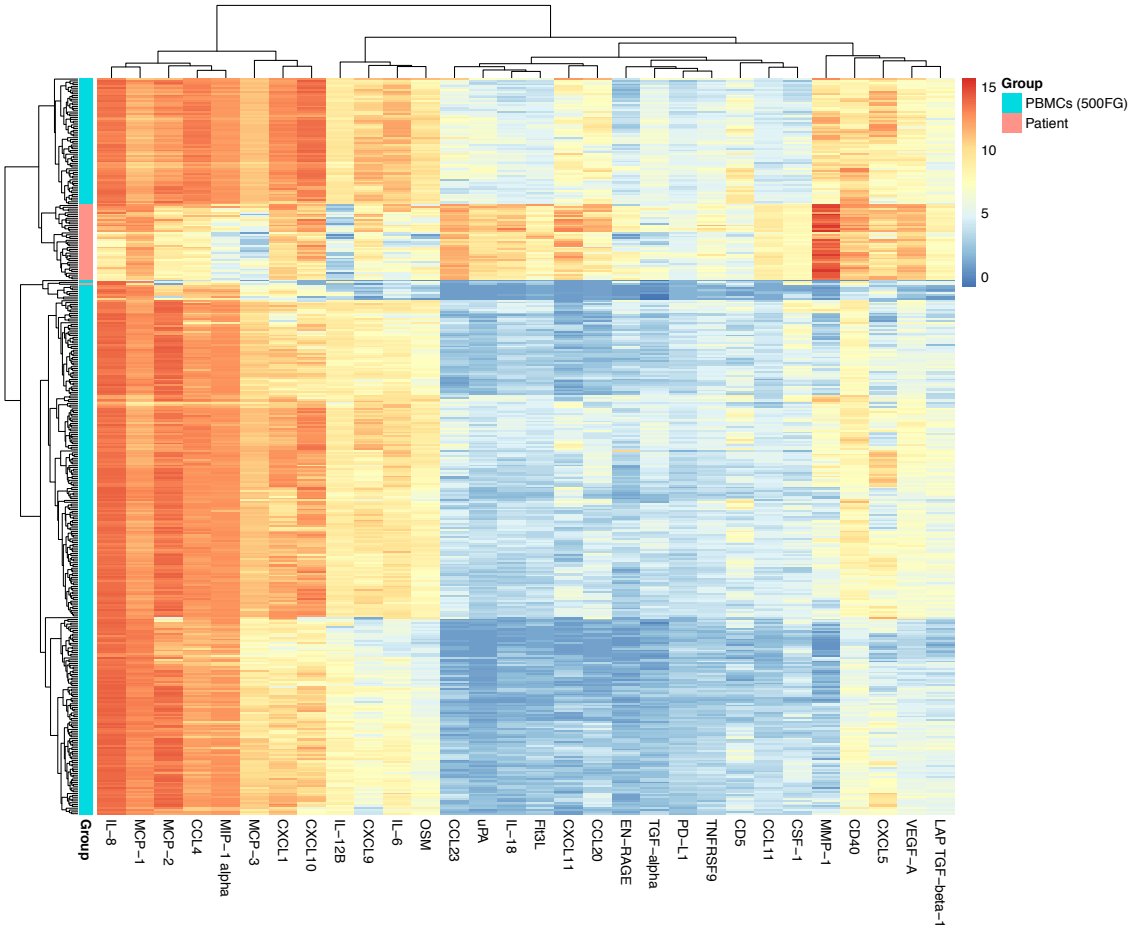



**Figure S4:** Unsupervised clustering of inflammatory proteins profiled in *Candida*-stimulated PBMCs and patients showed different clustering patterns between these two groups, suggesting that additional factors in plasma of patients contribute to inflammatory responses in patients. Proteins were detected in at least 90% of the samples and were expressed in both patients and PBMC samples ($N_{proteins} = 30$).

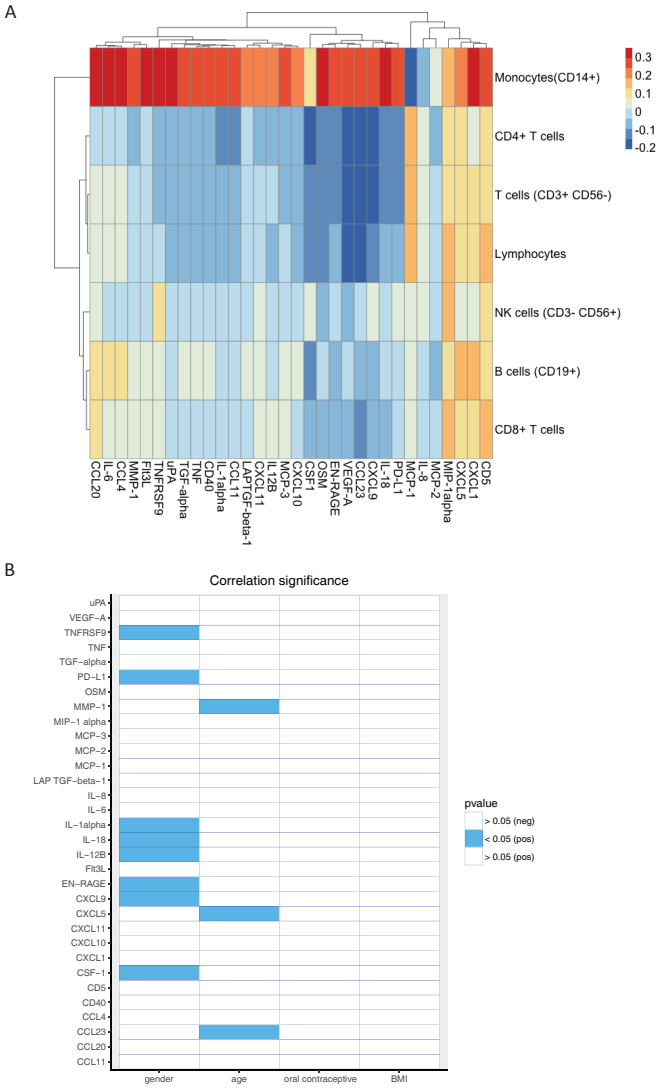



**Figure S5:** Minimal effect of cellc counts and non-genetic host factors on inflammatory proteins in response to *C. albicans*. (A) Cell counts weakly influence the levels of inflammatory proteins in PBMCs stimulated with *C. albicans* yeast for 24 hours. Unsupervised hierarchical clustering was performed using Spearman correlation as the measure of similarity. The red color depicts the strong positive correlation whereas blue color indicates the strong negative correlation. (B) Correlations of age, gender, oral contraceptive use and BMI to protein production released from PBMCs upon stimulation with *C. albicans* yeast. Significance of non-genetic factors (x axis) in relation to different proteins (y axis) is shown. P values have been FDR-corrected. Blue indicates a significant positive correlation (P value < 0.05). All significantly correlated proteins with age or gender showed a positive (pos) correlation.

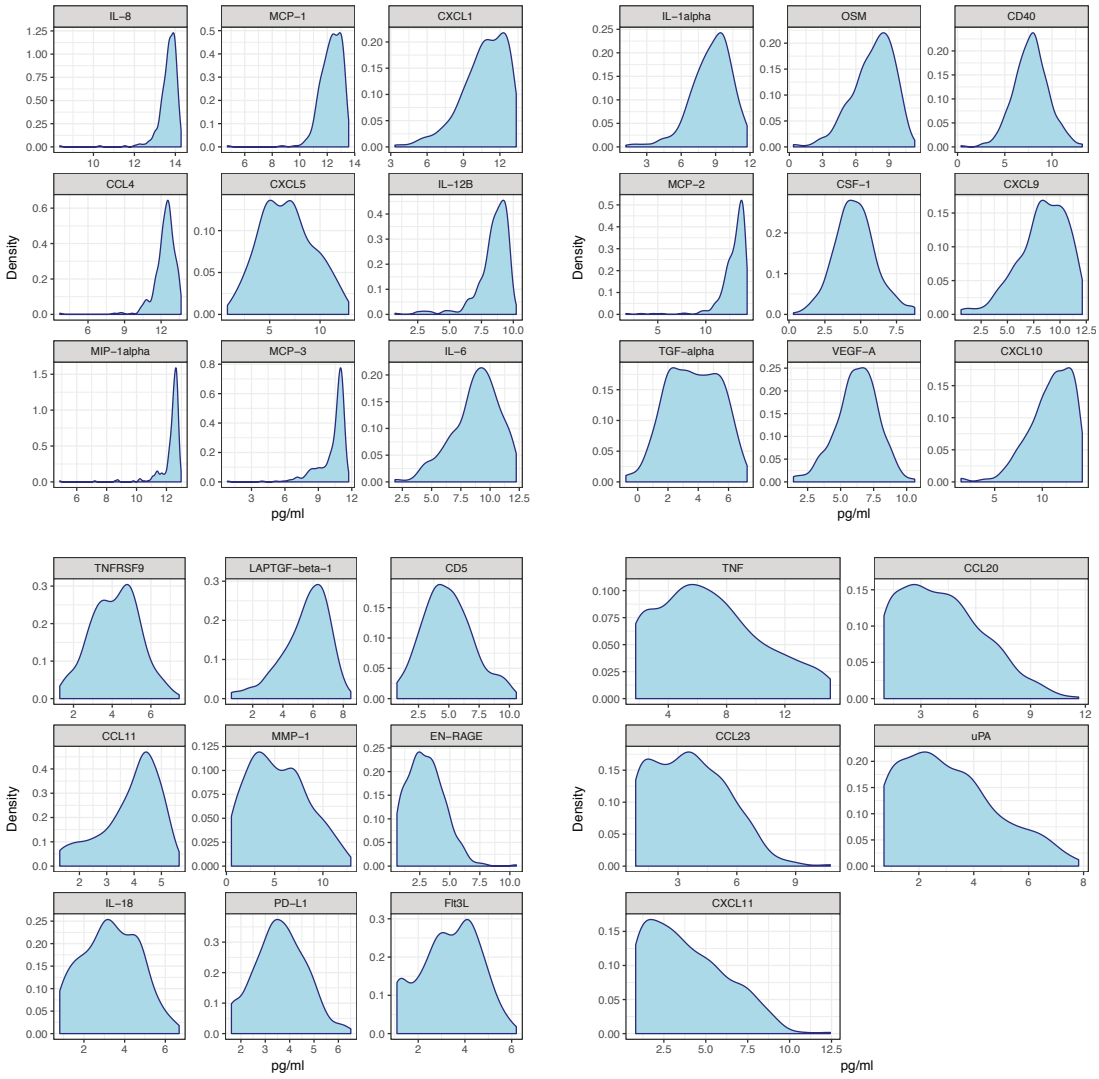



**Figure S6:** Density plots of the distribution of inflammatory proteins released from PBMCs stimulated with *C. albicans* yeast isolated from healthy individuals of 500FG cohort. Majority of proteins follows non-normal distribution after log2 transformation, with the exception of three proteins, CD40, VEGF-A and TNFRSF9. Related to Table S3.

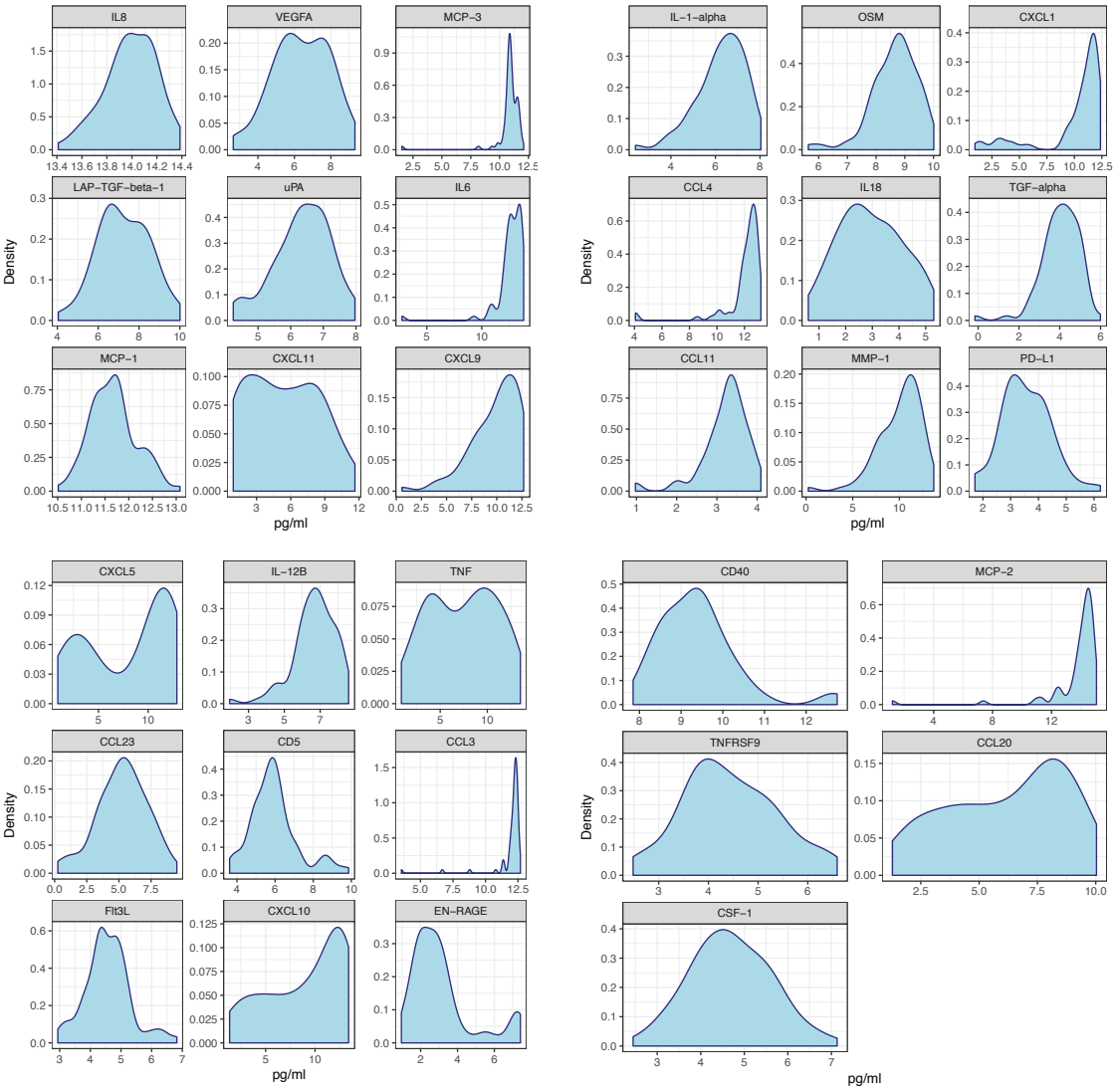
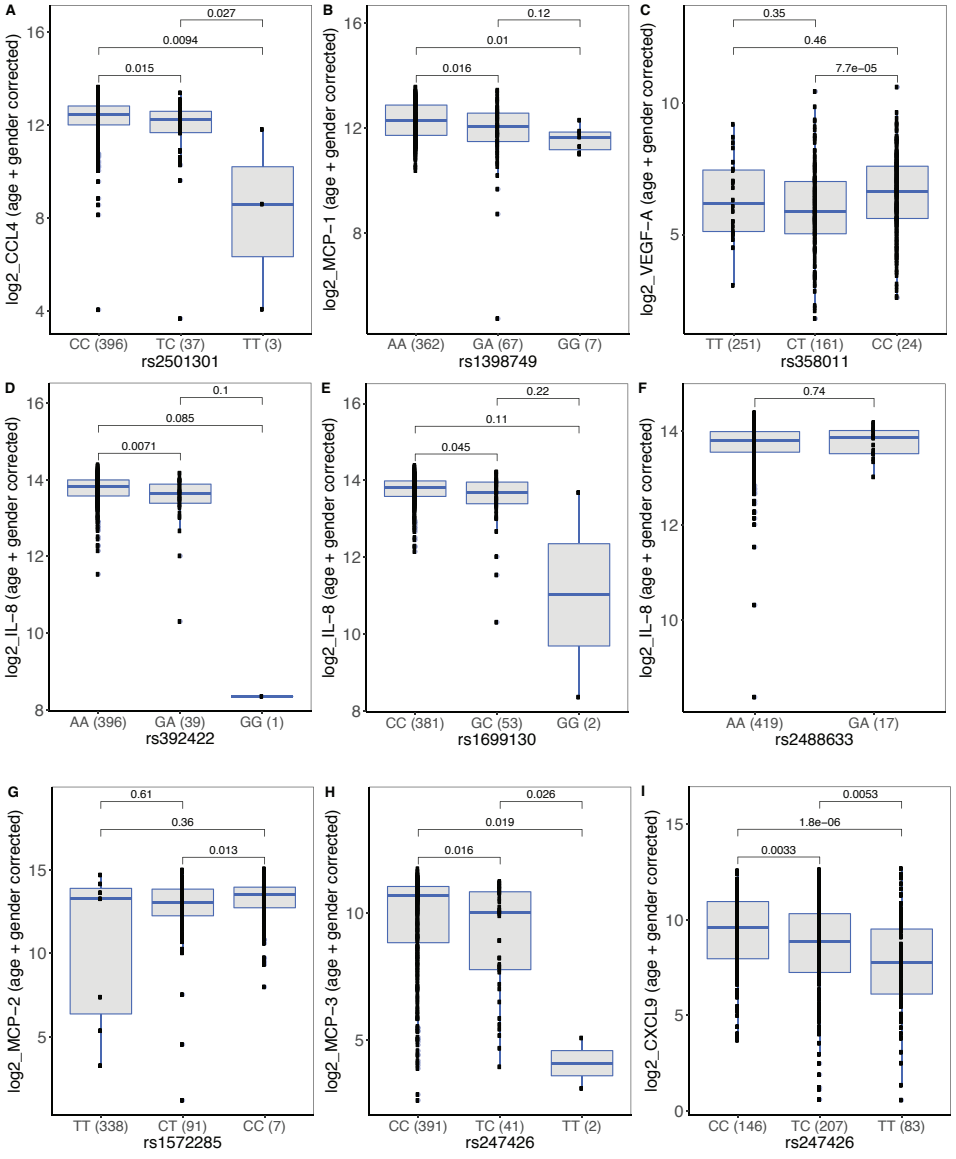
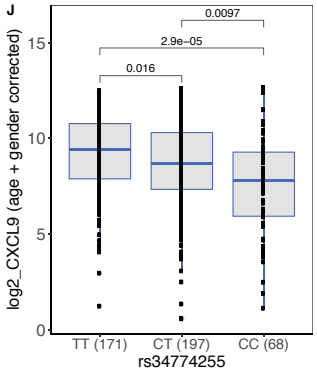



**Figure S7:** Density plots of the distribution of inflammatory proteins released from PBMCs stimulated with *C. albicans* yeast isolated from healthy individuals of Lifelines cohort. Majority of proteins follows non-normal distribution after log2 transformation. Related to Table S7.

**Figure S8:** Box plots showing the association of genotypes from genome-wide significant pQTLs identified in the joint analysis between the 500FG and Lifelines DEEP cohorts with *Candida*-induced protein levels. Boxplot of (A) SNP rs2501301 with CCL4 levels, (B) SNP rs1398749 with MCP-1 levels, (C) SNP rs358011 with VEGF-A levels, (D) SNP rs392422 with IL-8 levels, (E) SNP rs1699130 with IL-8 levels, (F) SNP rs2488633 with IL-8, (G) SNP rs1572285 with MCP-2, (H) SNP rs247426 with MCP-3, (I) SNP rs247426 with CXCL9, and (J) rs34774255 with CXCL9 levels. The numbers of individuals per genotype are shown in parentheses below each box plot. P values were obtained using Wilcoxon Rank-Sum test to compare if there are statistically significant differences of protein levels between SNP genotypes. $P < 0.05$ was considered to be the threshold of significance. Related to Table 1.





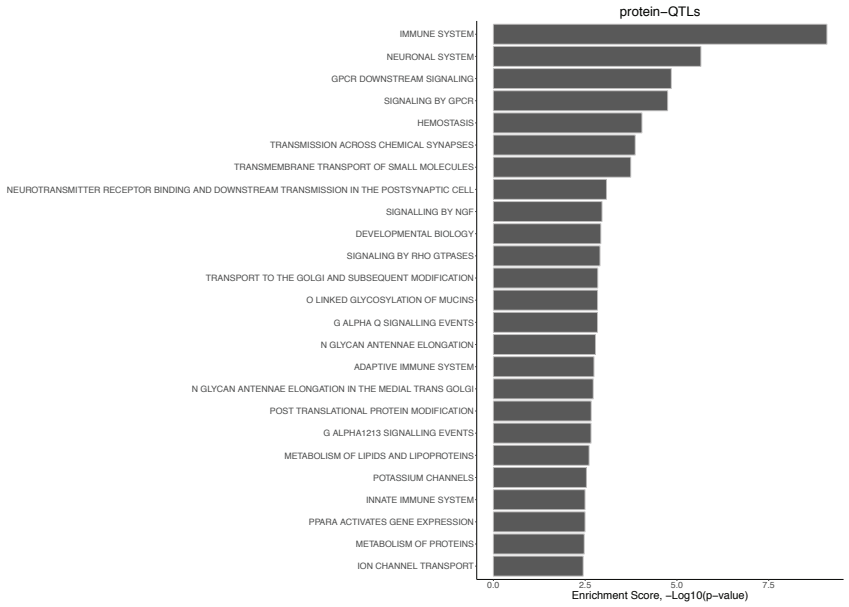




**Figure S9:** Pathway enrichment analysis of *Candida*-induced pQTLs with a P value < $9.99 \times 10^{-4}$ showed an over-representation of genes involved in hemostasis and metabolism of lipids and lipoproteins based on Reactome database. x axis shows the enrichment p-values for each pathway.

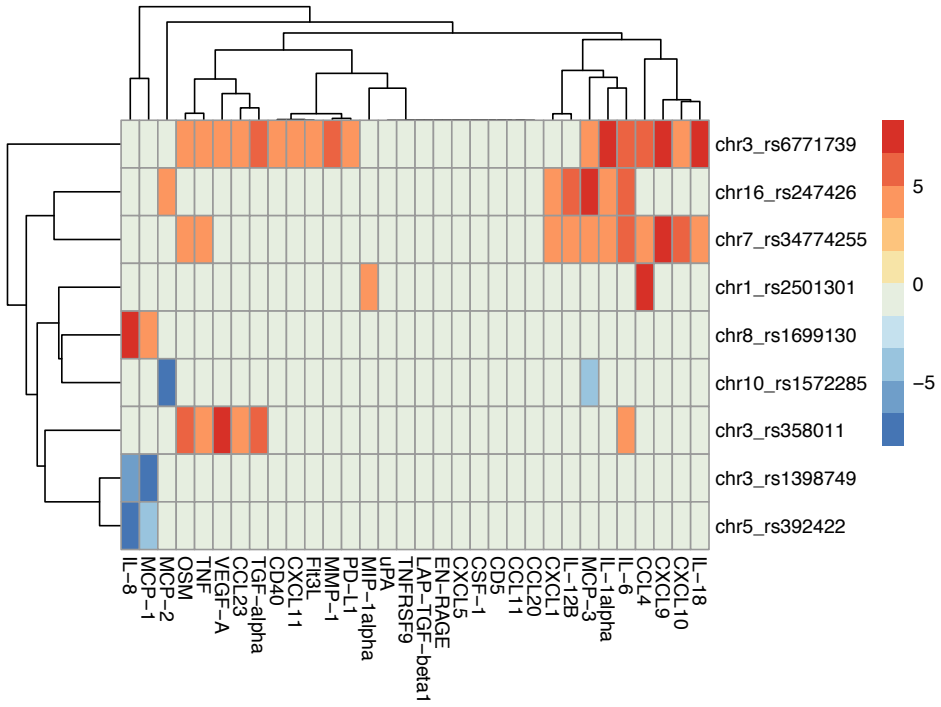



**Figure S10:** Heatmap shows all genome-wide significant pQLTs that have at least two associations with distinct proteins at P threshold $1.56x10^{-4}$ when correcting for multiple testing. All associations show the same allelic direction in both 500FG and Lifelines DEEP cohorts ($P_{heterogeneity} \geq 0.05$). Associations with P value above this cut-off are set zero for visualization purposes. For clustering, we used an unsupervised clustering approach based on the negative log10 of the P values. Color scale indicates the direction of effect, with red showing a positive z score value and blue showing a negative z score value. Related to Fig. 4.


**Figure S11:** (A) Multidimensional scaling analysis (MDS) to check for population stratification in the candidaemia cohort consisted of 178 cases. Out of 178 cases, 65 were non-survivors and 83 survivors. We do not have available mortality data for the rest of cases. Dimensions were calculated on the basis of the identity-by-descent (IBD) pairwise distance among all patients and the first–dimension values were plotted against the second-dimension values. Non-survivors (n = 65), survivors (n = 83) and the rest of candidaemia cases are highlighted in different colors. (B) Quantile-quantile (QQ) plot of log10 (P values) of primary GWAS P-values to those expected for a null distribution. The genetic inflation factor λ for all SNPs was 1, indicating no population stratification.

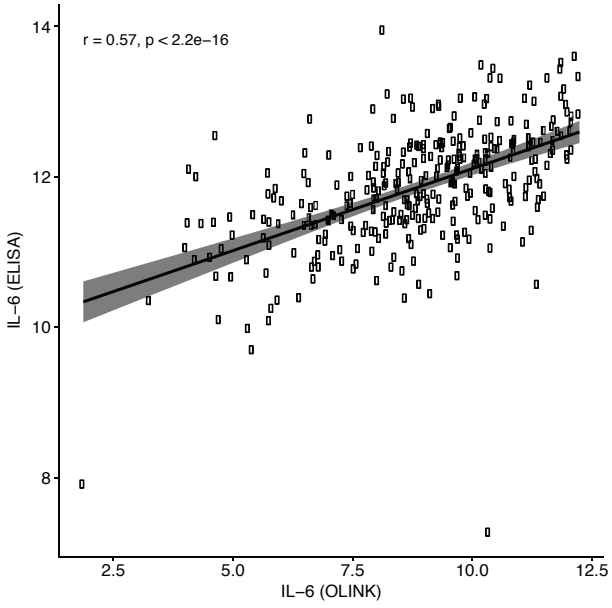



**Figure S12:** A positive Spearman correlation coefficient was observed between protein concentrations of IL-6 in log2 picogram/ml measured with ELISA and NPX values as measured with OLINK technology.

**Table S1.** Inflammatory proteins measured with high-throughput OLINK technology in plasma samples of candidaemia patients. Out of 92 proteins measured, 64 were detected in at least 90% of the samples.

| Protein | Uniprot no. | Gene | Ensembl_id | Chromosome _gene | Start_gene | End_gene |
|---|---|---|---|---|---|---|
| 4E-BP1 | Q13541 | *EIF4EBP1* | ENSG00000187840 | 8 | 37888020 | 37917883 |
| ADA | P00813 | *ADA* | ENSG00000196839 | 20 | 43248163 | 43280376 |
| CASP-8 | Q14790 | *CASP8* | ENSG00000064012 | 2 | 202122754 | 202152434 |
| CCL11 | P51671 | *CCL11* | ENSG00000172156 | 17 | 32612687 | 32615199 |
| CCL19 | Q99731 | *CCL19* | ENSG00000172724 | 9 | 34689567 | 34691274 |
| CCL20 | P78556 | *CCL20* | ENSG00000115009 | 2 | 228678558 | 228682280 |
| CCL23 | P55773 | *CCL23* | ENSG00000274736 | 17 | 34340097 | 34345005 |
| CCL25 | O15444 | *CCL25* | ENSG00000131142 | 19 | 8052767 | 8062650 |
| CCL4 | P13236 | *CCL4* | ENSG00000275302 | 17 | 34431220 | 34433014 |
| CD244 | Q9BZW8 | *CD244* | ENSG00000122223 | 1 | 160799950 | 160832692 |
| CD40 | P25942 | *CD40* | ENSG00000101017 | 20 | 44746906 | 44758384 |
| CD5 | P06127 | *CD5* | ENSG00000110448 | 11 | 60869930 | 60895323 |
| CD6 | Q8WWJ7 | *CD6* | ENSG00000013725 | 11 | 60739113 | 60787848 |
| CDCP1 | Q9H5V8 | *CDCP1* | ENSG00000163814 | 3 | 45123769 | 45187914 |
| CSF-1 | P09603 | *CSF1* | ENSG00000184371 | 1 | 110453233 | 110473616 |
| CST5 | P28325 | *CST5* | ENSG00000170367 | 20 | 23856572 | 23860380 |
| CX3CL1 | P78423 | *CX3CL1* | ENSG00000006210 | 16 | 57372458 | 57385048 |
| CXCL1 | P09341 | *CXCL1* | ENSG00000163739 | 4 | 74735109 | 74737019 |
| CXCL10 | P02778 | *CXCL10* | ENSG00000169245 | 4 | 76942269 | 76944689 |
| CXCL11 | O14625 | *CXCL11* | ENSG00000169248 | 4 | 76954840 | 76957350 |
| CXCL5 | P42830 | *CXCL5* | ENSG00000163735 | 4 | 74861359 | 74864446 |
| CXCL6 | P80162 | *CXCL6* | ENSG00000124875 | 4 | 74702273 | 74704477 |
| CXCL9 | Q07325 | *CXCL9* | ENSG00000138755 | 4 | 76922623 | 76928641 |
| DNER | Q8NFT8 | *DNER* | ENSG00000187957 | 2 | 229357629 | 229714558 |
| EN-RAGE | P80511 | *S100A12* | ENSG00000163221 | 1 | 153346184 | 153348075 |
| FGF-19 | O95750 | *FGF19* | ENSG00000162344 | 11 | 69698232 | 69704642 |
| FGF-21 | Q9NSA1 | *FGF21* | ENSG00000105550 | 19 | 48755559 | 48758330 |
| FGF-23 | Q9GZV9 | *FGF23* | ENSG00000118972 | 12 | 4368227 | 4379728 |
| Flt3L | P49771 | *FLT3LG* | ENSG00000090554 | 19 | 49977818 | 49990894 |
| HGF | P14210 | *HGF* | ENSG00000019991 | 7 | 81331444 | 81399452 |
| IL-10 | P22301 | *IL10* | ENSG00000136634 | 1 | 206940948 | 206945839 |
| IL-10RB | Q08334 | *IL10RB* | ENSG00000243646 | 21 | 33266358 | 33,310,187 |
| IL-12B | P29460 | *IL12B* | ENSG00000113302 | 5 | 158741791 | 158757481 |
| IL-18 | Q14116 | *IL18* | ENSG00000150782 | 11 | 112013974 | 112034840 |
| IL-18R1 | Q13478 | *IL18R1* | ENSG00000115604 | 2 | 102979093 | 103015217 |
| IL-6 | P05231 | *IL6* | ENSG00000136244 | 7 | 22766766 | 22771621 |
| IL-7 | P13232 | *IL7* | ENSG00000104432 | 8 | 78675743 | 78805523 |
| IL-8 | P10145 | *CXCL8* | ENSG00000169429 | 4 | 74606223 | 74609433 |
| LAP TGF-beta-1 | P01137 | *TGFB1* | ENSG00000105329 | 19 | 41836812 | 41859831 |
| LIF-R | P42702 | *LIFR* | ENSG00000113594 | 5 | 38474963 | 38608354 |
| MCP-1 | P13500 | *CCL2* | ENSG00000108691 | 17 | 32582296 | 32584220 |
| MCP-2 | P80075 | *CCL8* | ENSG00000108700 | 17 | 32646066 | 32648421 |
| MCP-3 | P80098 | *CCL7* | ENSG00000108688 | 17 | 32597235 | 32599261 |
| MCP-4 | Q99616 | *CCL13* | ENSG00000181374 | 17 | 32683471 | 32685629 |
| MIP-1 alpha | P10147 | *CCL3* | ENSG00000277632 | 17 | 34415603 | 34417506 |
| MMP-1 | P03956 | *MMP1* | ENSG00000196611 | 11 | 102660641 | 102668966 |
| MMP-10 | P09238 | *MMP10* | ENSG00000166670 | 11 | 102641233 | 102651359 |
| OPG | O00300 | *TNFRSF11B* | ENSG00000164761 | 8 | 119935796 | 119964383 |
| OSM | P13725 | *OSM* | ENSG00000099985 | 22 | 30658819 | 30662829 |
| PD-L1 | Q9NZQ7 | *CD274* | ENSG00000120217 | 9 | 5450503 | 5470567 |
| SCF | P21583 | *KITLG* | ENSG00000049130 | 12 | 88492793 | 88580851 |
| SIRT2 | Q8IXJ6 | *SIRT2* | ENSG00000068903 | 19 | 39369195 | 39390502 |
| SLAMF1 | Q13291 | *SLAMF1* | ENSG00000117090 | 1 | 160608100 | 160647295 |
| ST1A1 | P50225 | *SULT1A1* | ENSG00000196502 | 16 | 28605196 | 28623625 |
| STAMPB | O95630 | *STAMBP* | ENSG00000124356 | 2 | 74056043 | 74094295 |

| Protein | Uniprot no. | Gene | Ensembl_id | Chromosome _gene | Start_gene | End_gene |
|---|---|---|---|---|---|---|
| TGF-alpha | P01135 | *TGFA* | ENSG00000163235 | 2 | 70674412 | 70781147 |
| TNFB | P01374 | *LTA* | ENSG00000226275 | 6 | 31539876 | 31542100 |
| TNFRSF9 | Q07011 | *TNFRSF9* | ENSG00000049249 | 1 | 7975931 | 8003225 |
| TNFSF14 | O43557 | *TNFSF14* | ENSG00000125735 | 19 | 6663148 | 6670599 |
| TRAIL | P50591 | *TNFSF10* | ENSG00000121858 | 3 | 172223298 | 172241297 |
| TRANCE | O14788 | *TNFSF11* | ENSG00000120659 | 13 | 43148291 | 43182149 |
| TWEAK | O43508 | *TNFSF12* | ENSG00000239697 | 17 | 7452375 | 7461207 |
| uPA | P00749 | *PLAU* | ENSG00000122861 | 10 | 75670862 | 75677258 |
| VEGF-A | P15692 | *VEGFA* | ENSG00000112715 | 6 | 43737946 | 43754223 |

**Table S2.** Fold change of expression of inflammatory proteins in plasma of candidaemia patients compared to baseline healthy plasma samples (Lifelines). Proteins were measured by the high-throughput OLINK technology using the inflammatory array.

| Protein | Fold change | Log2 Fold change | Protein | Fold change | Log2 Fold-change |
|---|---|---|---|---|---|
| CASP-8 | 4,540 | 2,183 | Flt3L | 0,957 | -0,063 |
| TGF-alpha | 4,185 | 2,065 | TWEAK | 0,950 | -0,074 |
| OSM | 4,122 | 2,044 | SCF | 0,929 | -0,106 |
| EN-RAGE | 3,292 | 1,719 | SLAMF1 | 0,924 | -0,114 |
| MCP-3 | 3,049 | 1,608 | ADA | 0,924 | -0,114 |
| MIP-1 alpha | 2,898 | 1,535 | CST5 | 0,910 | -0,137 |
| TNFSF14 | 2,437 | 1,285 | FGF-19 | 0,898 | -0,156 |
| IL-6 | 2,177 | 1,122 | IL-7 | 0,853 | -0,229 |
| MCP-4 | 2,149 | 1,104 | CD244 | 0,820 | -0,287 |
| FGF-21 | 1,910 | 0,933 | CCL25 | 0,817 | -0,291 |
| MMP-1 | 1,904 | 0,929 | MMP-10 | 0,796 | -0,330 |
| IL-8 | 1,820 | 0,864 | IL-12B | 0,793 | -0,334 |
| CXCL11 | 1,722 | 0,784 | TRAIL | 0,778 | -0,362 |
| ST1A1 | 1,712 | 0,775 | TNFB | 0,687 | -0,542 |
| CCL4 | 1,614 | 0,691 | TRANCE | 0,510 | -0,972 |
| IL-18 | 1,497 | 0,582 | CD6 | 1,106 | 0,146 |
| CXCL1 | 1,484 | 0,570 | IL-10 | 1,072 | 0,100 |
| CCL20 | 1,480 | 0,565 | TNFRSF9 | 1,059 | 0,083 |
| CD5 | 1,459 | 0,545 | MCP-2 | 1,043 | 0,060 |
| LIF-R | 1,441 | 0,527 | CX3CL1 | 1,036 | 0,051 |
| SIRT2 | 1,430 | 0,516 | uPA | 1,007 | 0,010 |
| CXCL6 | 1,419 | 0,505 | CSF-1 | 1,004 | 0,006 |
| CDCP1 | 1,412 | 0,498 | CXCL5 | 0,992 | -0,011 |
| FGF-23 | 1,410 | 0,495 | | | |
| HGF | 1,400 | 0,485 | | | |
| STAMPB | 1,395 | 0,480 | | | |
| 4E-BP1 | 1,371 | 0,455 | | | |
| MCP-1 | 1,326 | 0,407 | | | |
| CD40 | 1,246 | 0,317 | | | |
| LAP TGF-beta-1 | 1,241 | 0,312 | | | |
| CCL19 | 1,228 | 0,296 | | | |
| CCL23 | 1,201 | 0,265 | | | |
| CXCL10 | 1,151 | 0,203 | | | |
| CCL11 | 1,144 | 0,194 | | | |
| VEGF-A | 1,128 | 0,174 | | | |
| IL-18R1 | 1,110 | 0,151 | | | |
| PD-L1 | 1,108 | 0,148 | | | |
| CXCL9 | 1,083 | 0,115 | | | |
| OPG | 1,063 | 0,089 | | | |
| IL-10RB | 1,039 | 0,055 | | | |
| DNER | 1,012 | 0,018 | | | |

**Table S3.** Inflammatory proteins measured with high-throughput OLINK techonology in supernatants of PBMCs isolated from individuals of 500FG cohort. We used proteins that were detected in at least 90% of the samples for further analysis. Shapiro test was used to test for normality under the null hypothesis that the protein distribution is normal. If P value < 0.05, the distribution is non-normal. Most proteins follow non-normal distribution. Chr: chromosome

| Protein | P value | Uniprot no. | Gene | Ensembl_id | Chr_gene | Start_gene | End_gene |
|---|---|---|---|---|---|---|---|
| CCL11 | 4,48E-11 | P51671 | *CCL11* | ENSG00000172156 | 17 | 32612687 | 32615199 |
| CCL20 | 1,79E-07 | P78556 | *CCL20* | ENSG00000115009 | 2 | 228678558 | 228682280 |
| CCL23 | 5,70E-06 | P55773 | *CCL23* | ENSG00000274736 | 17 | 34340097 | 34345005 |
| CCL4 | 6,07E-21 | P13236 | *CCL4* | ENSG00000275302 | 17 | 34431220 | 34433014 |
| CD40 | 6,65E-01 | P25942 | *CD40* | ENSG00000101017 | 20 | 44746906 | 44758384 |
| CD5 | 2,39E-04 | P06127 | *CD5* | ENSG00000110448 | 11 | 60869930 | 60895323 |
| CSF-1 | 2,61E-02 | P09603 | *CSF1* | ENSG00000184371 | 1 | 110453233 | 110473616 |
| CXCL1 | 1,53E-11 | P09341 | *CXCL1* | ENSG00000163739 | 4 | 74735109 | 74737019 |
| CXCL10 | 9,79E-10 | P02778 | *CXCL10* | ENSG00000169245 | 4 | 76942269 | 76944689 |
| CXCL11 | 7,29E-09 | O14625 | *CXCL11* | ENSG00000169248 | 4 | 76954840 | 76957350 |
| CXCL5 | 1,43E-03 | P42830 | *CXCL5* | ENSG00000163735 | 4 | 74861359 | 74864446 |
| CXCL9 | 9,28E-08 | Q07325 | *CXCL9* | ENSG00000138755 | 4 | 76922623 | 76928641 |
| EN-RAGE | 1,50E-07 | P80511 | *S100A12* | ENSG00000163221 | 1 | 153346184 | 153348075 |
| Flt3L | 4,01E-03 | P49771 | *FLT3LG* | ENSG00000090554 | 19 | 49977818 | 49990894 |
| IL-12B | 5,31E-21 | P29460 | *IL12B* | ENSG00000113302 | 5 | 158741791 | 158757481 |
| IL-18 | 4,28E-04 | Q14116 | *IL18* | ENSG00000150782 | 11 | 112013974 | 112034840 |
| IL-1alpha | 7,15E-08 | P01583 | *IL1A* | ENSG00000115008 | 2 | 113531492 | 113542971 |
| IL-6 | 1,67E-06 | P05231 | *IL6* | ENSG00000136244 | 7 | 22766766 | 22771621 |
| IL-8 | 3,13E-25 | P10145 | *CXCL8* | ENSG00000169429 | 4 | 74606223 | 74609433 |
| LAP TGF-beta-1 | 2,15E-08 | P01137 | *TGFB1* | ENSG00000105329 | 19 | 41836812 | 41859831 |
| MCP-1 | 4,22E-18 | P13500 | *CCL2* | ENSG00000108691 | 17 | 32582296 | 32584220 |
| MCP-2 | 1,11E-23 | P80075 | *CCL8* | ENSG00000108700 | 17 | 32646066 | 32648421 |
| MCP-3 | 1,71E-21 | P80098 | *CCL7* | ENSG00000108688 | 17 | 32597235 | 32599261 |
| MIP-1 alpha | 6,23E-29 | P10147 | *CCL3* | ENSG00000277632 | 17 | 34415603 | 34417506 |
| MMP-1 | 1,18E-06 | P03956 | *MMP1* | ENSG00000196611 | 11 | 102660641 | 102668966 |
| OSM | 6,56E-07 | P13725 | *OSM* | ENSG00000099985 | 22 | 30658819 | 30662829 |
| PD-L1 | 2,78E-02 | Q9NZQ7 | *CD274* | ENSG00000120217 | 9 | 5450503 | 5470567 |
| TGF-alpha | 9,00E-05 | P01135 | *TGFA* | ENSG00000163235 | 2 | 70674412 | 70781147 |
| TNF | 1,03E-07 | P01375 | *TNF* | ENSG00000228321 | 6 | 31544292 | 31546112 |
| TNFRSF9 | 1,05E-01 | Q07011 | *TNFRSF9* | ENSG00000049249 | 1 | 7975931 | 8003225 |
| uPA | 1,09E-08 | P00749 | *PLAU* | ENSG00000122861 | 10 | 75670862 | 75677258 |
| VEGF-A | 3,00E-01 | P15692 | *VEGFA* | ENSG00000112715 | 6 | 43737946 | 43754223 |

**Table S4.** Fold change of expression of inflammatory proteins in supernatant samples of PBMCs stimulated with *C. albicans* yeast compared to supernatant samples stimulated with RPMI medium. Samples were stimulated either with C. albicans yeast or RPMI for 24 hours and proteins were measured with the inflammatory array of the high-throughput OLINK technology.

| Protein | Fold_change | Log2_Fold_change |
|---|---|---|
| IL-12B | 19,15 | 4,26 |
| CXCL9 | 9,76 | 3,29 |
| MCP-2 | 7,63 | 2,93 |
| IL-1alpha | 6,68 | 2,74 |
| MIP-1 alpha | 6,51 | 2,70 |
| MCP-3 | 6,43 | 2,69 |
| TGF-alpha | 5,78 | 2,53 |
| IL-6 | 4,71 | 2,24 |
| CCL23 | 4,25 | 2,09 |
| CCL20 | 4,21 | 2,07 |
| CXCL11 | 4,16 | 2,06 |
| IL-18 | 4,09 | 2,03 |
| TNF | 3,88 | 1,96 |
| CXCL1 | 3,87 | 1,95 |
| CXCL10 | 3,74 | 1,90 |
| CCL4 | 3,69 | 1,88 |
| CSF-1 | 3,39 | 1,76 |
| CXCL5 | 3,13 | 1,65 |
| uPA | 3,07 | 1,62 |
| CCL11 | 3,05 | 1,61 |
| MMP-1 | 2,62 | 1,39 |
| EN-RAGE | 2,47 | 1,30 |
| MCP-1 | 2,35 | 1,23 |
| PD-L1 | 2,25 | 1,17 |
| OSM | 2,23 | 1,16 |
| TNFRSF9 | 1,83 | 0,87 |
| Flt3L | 1,66 | 0,73 |
| VEGF-A | 1,57 | 0,65 |
| IL-8 | 1,56 | 0,64 |
| CD40 | 1,38 | 0,46 |
| LAP TGF-beta-1 | 1,38 | 0,46 |
| CD5 | 1,35 | 0,43 |

**Table S5.** Correlations of gender, age, oral contraceptive use and BMI with inflammatory proteins released from PBMCs upon stimulation with *C. albicans* yeast for 24 hours. Inflammatory proteins were detected in at least 90% of the samples. Correlations calculated using a rank-based regression model, as described in Materials and Methods. The P values were FDR corrected. Proteins measured with the inflammatory array of the high-throughput OLINK technology.

| Protein | Gender | Age | Oral_contraceptive | BMI |
|---|---|---|---|---|
| IL-8 | -9,77E-01 | -5,30E-01 | -8,85E-01 | 4,60E-01 |
| CXCL1 | -6,78E-01 | 1,64E-01 | 4,80E-01 | -7,87E-01 |
| MCP-1 | -1,46E-01 | -3,76E-01 | 6,20E-01 | 4,48E-01 |
| CXCL5 | -5,47E-02 | 4,72E-02 | 6,81E-02 | -1,37E-01 |
| TNFRSF9 | 9,92E-04 | 8,56E-01 | -2,30E-01 | -9,12E-01 |
| IL-18 | 1,15E-03 | 1,14E-01 | -1,20E-01 | -8,69E-01 |
| CXCL9 | 1,50E-03 | 4,60E-01 | -1,58E-01 | -7,77E-01 |
| EN-RAGE | 9,90E-03 | 1,37E-01 | -2,98E-01 | -5,72E-01 |
| CSF-1 | 9,90E-03 | 4,09E-01 | -7,63E-01 | -9,77E-01 |
| PD-L1 | 1,62E-02 | 4,01E-01 | -2,38E-01 | 9,24E-01 |
| IL-12B | 1,62E-02 | 8,97E-01 | -9,08E-01 | 9,08E-01 |
| IL-1alpha | 2,55E-02 | 5,26E-01 | -5,58E-01 | -8,42E-01 |
| TNF | 6,01E-02 | 4,60E-01 | -4,48E-01 | -5,96E-01 |
| MCP-3 | 1,11E-01 | 2,38E-01 | -1,00E+00 | -4,57E-01 |
| CCL23 | 1,14E-01 | 9,82E-03 | -5,72E-01 | -4,09E-01 |
| CXCL11 | 1,59E-01 | 3,45E-01 | -8,56E-01 | -8,56E-01 |
| CCL11 | 1,73E-01 | 2,99E-01 | -9,08E-01 | -5,35E-01 |
| uPA | 1,76E-01 | 1,54E-01 | -3,22E-01 | -3,30E-01 |
| CD40 | 2,51E-01 | 3,07E-01 | -4,27E-01 | -5,99E-01 |
| CXCL10 | 2,62E-01 | 2,83E-01 | -5,99E-01 | -5,63E-01 |
| OSM | 3,07E-01 | 6,67E-02 | -6,49E-01 | -2,51E-01 |
| CD5 | 3,69E-01 | 5,57E-01 | -1,00E+00 | -6,00E-01 |
| Flt3L | 3,76E-01 | 4,98E-01 | -4,98E-01 | -6,78E-01 |
| MIP-1 alpha | 4,39E-01 | 6,81E-01 | 6,49E-01 | -8,69E-01 |
| CCL4 | 4,57E-01 | 6,27E-01 | -9,30E-01 | -5,96E-01 |
| TGF-alpha | 4,60E-01 | 3,30E-01 | -4,60E-01 | -7,77E-01 |
| LAP TGF-beta-1 | 8,34E-01 | 6,38E-02 | -6,82E-01 | -1,37E-01 |
| MMP-1 | 8,42E-01 | 9,90E-03 | -9,77E-01 | -2,63E-01 |
| MCP-2 | 8,56E-01 | -9,77E-01 | 5,90E-01 | -9,08E-01 |
| VEGF-A | 8,56E-01 | 6,70E-02 | -8,56E-01 | -4,60E-01 |
| IL-6 | 8,96E-01 | 1,73E-01 | 7,42E-01 | -4,60E-01 |
| CCL20 | 9,77E-01 | 6,01E-02 | 9,77E-01 | -4,57E-01 |

**Table S6.** Correlations of circulating mediators, resistin, adiponectin, and alpha-1 Antitrypsin (AAT), leptin and CRP, with inflammatory proteins released from PBMCs upon stimulation with *C. albicans* yeast for 24 hours. Inflammatory proteins were detected in at least 90% of the samples. Correlations calculated using a rank-based regression model, as described in Materials and Methods. The P values were FDR corrected.

| Protein | Resistin | Adiponectin | AAT | Leptin | CRP |
|---|---|---|---|---|---|
| CCL11 | -9,97E-01 | 1,00E+00 | -1,00E+00 | -9,27E-01 | 1,00E+00 |
| CCL20 | -1,00E+00 | -9,09E-01 | 1,00E+00 | -9,09E-01 | -1,00E+00 |
| CCL23 | -1,00E+00 | -1,00E+00 | 9,68E-01 | -9,27E-01 | -1,00E+00 |
| CCL4 | -9,14E-01 | -9,14E-01 | 9,68E-01 | 9,14E-01 | 1,00E+00 |
| CD40 | 9,68E-01 | -8,70E-01 | 9,68E-01 | -9,14E-01 | -1,00E+00 |
| CD5 | 9,14E-01 | -9,14E-01 | 9,14E-01 | 1,00E+00 | 1,00E+00 |
| CSF-1 | 1,00E+00 | -1,00E+00 | 9,68E-01 | -1,00E+00 | 1,00E+00 |
| CXCL1 | -9,56E-01 | -9,68E-01 | 1,00E+00 | -9,09E-01 | 1,00E+00 |
| CXCL10 | -8,70E-01 | -9,14E-01 | 9,14E-01 | 9,87E-01 | -9,97E-01 |
| CXCL11 | -5,55E-01 | -9,02E-01 | 9,14E-01 | 9,68E-01 | -9,89E-01 |
| CXCL5 | 1,00E+00 | -1,00E+00 | -9,68E-01 | -9,14E-01 | 1,00E+00 |
| CXCL9 | -9,27E-01 | -9,25E-01 | 9,09E-01 | 1,00E+00 | -1,00E+00 |
| EN-RAGE | 1,00E+00 | -9,68E-01 | 9,68E-01 | -9,09E-01 | -9,14E-01 |
| Flt3L | -9,68E-01 | -8,70E-01 | 9,27E-01 | -1,00E+00 | -1,00E+00 |
| IL-12B | 9,68E-01 | -8,70E-01 | 9,97E-01 | 9,97E-01 | 9,87E-01 |
| IL-18 | -9,27E-01 | -9,81E-01 | 8,70E-01 | 1,00E+00 | -9,68E-01 |
| IL-1alpha | -9,56E-01 | -9,14E-01 | 9,14E-01 | -9,68E-01 | -9,27E-01 |
| IL-6 | -9,14E-01 | -9,14E-01 | 9,14E-01 | -9,68E-01 | 9,27E-01 |
| IL-8 | 1,00E+00 | 9,81E-01 | -1,00E+00 | -1,00E+00 | 9,68E-01 |
| LAP TGF-beta-1 | -1,00E+00 | -1,00E+00 | -9,28E-01 | -8,70E-01 | -9,68E-01 |
| MCP-1 | 1,00E+00 | 9,14E-01 | -9,74E-01 | 1,00E+00 | 1,00E+00 |
| MCP-2 | 9,88E-01 | 9,09E-01 | -9,35E-01 | -9,85E-01 | 8,70E-01 |
| MCP-3 | -9,09E-01 | 1,00E+00 | 9,51E-01 | -9,56E-01 | 9,97E-01 |
| MIP-1 alpha | -1,00E+00 | 9,68E-01 | -8,70E-01 | -8,70E-01 | 9,56E-01 |
| MMP-1 | -9,14E-01 | -8,70E-01 | 9,27E-01 | -8,70E-01 | -1,00E+00 |
| OSM | -1,00E+00 | -9,14E-01 | 9,14E-01 | -9,02E-01 | -9,68E-01 |
| PD-L1 | -9,68E-01 | -9,56E-01 | 9,14E-01 | 9,28E-01 | -9,14E-01 |
| TGF-alpha | -1,00E+00 | -9,14E-01 | 9,09E-01 | -9,14E-01 | -9,27E-01 |
| TNF | -9,14E-01 | -8,70E-01 | 1,00E+00 | -1,00E+00 | 1,00E+00 |
| TNFRSF9 | -1,00E+00 | -9,14E-01 | 9,14E-01 | 1,00E+00 | -9,37E-01 |
| uPA | -9,88E-01 | -9,02E-01 | 9,68E-01 | -9,09E-01 | -9,88E-01 |
| VEGF-A | 1,00E+00 | -9,09E-01 | 9,14E-01 | -9,09E-01 | -9,56E-01 |

**Table S7.** Inflammatory proteins measured with high-throughput OLINK technonology in *Candida*-stimulated supernatants of PBMCs isolated from individuals of Lifelines DEEP cohort. We used proteins that were detected in at least 90% of the samples for further analysis. Shapiro test was used to test for normality under the null hypothesis that the protein distribution is normal. If P value < 0.05, the distribution is non-normal. Most proteins follow non-normal distribution. Chr: chromosome.

| Protein | P value | Uniprot no. | Gene | Ensembl_id | Chr_gene | Start_gene | End_gene |
|---|---|---|---|---|---|---|---|
| IL-8 | 2,45E-01 | P10145 | *CXCL8* | ENSG00000169429 | 4 | 74606223 | 74609433 |
| VEGF-A | 3,98E-01 | P15692 | *VEGFA* | ENSG00000112715 | 6 | 43737946 | 43754223 |
| MCP-3 | 2,28E-15 | P80098 | *CCL7* | ENSG00000108688 | 17 | 32597235 | 32599261 |
| CDCP1 | 1,03E-06 | Q9H5V8 | *CDCP1* | ENSG00000163814 | 3 | 45123769 | 45187914 |
| CD244 | 7,76E-07 | Q9BZW8 | *CD244* | ENSG00000122223 | 1 | 160799950 | 160832692 |
| OPG | 6,77E-01 | O00300 | *TNFRSF11B* | ENSG00000164761 | 8 | 119.935.796 | 119.964.383 |
| LAP TGF-beta-1 | 9,01E-01 | P01137 | *TGFB1* | ENSG00000105329 | 19 | 41836812 | 41859831 |
| uPA | 8,45E-02 | P00749 | *PLAU* | ENSG00000122861 | 10 | 75670862 | 75677258 |
| IL-6 | 2,00E-13 | P05231 | *IL6* | ENSG00000136244 | 7 | 22.766.766 | 22.771.621 |
| MCP-1 | 3,32E-01 | P13500 | *CCL2* | ENSG00000108691 | 17 | 32582296 | 32584220 |
| CXCL11 | 2,99E-03 | O14625 | *CXCL11* | ENSG00000169248 | 4 | 76954840 | 76957350 |
| CXCL9 | 1,14E-05 | Q07325 | *CXCL9* | ENSG00000138755 | 4 | 76922623 | 76928641 |
| IL-1alpha | 6,08E-03 | P01583 | *IL1A* | ENSG00000115008 | 2 | 113531492 | 113542971 |
| OSM | 5,20E-04 | P13725 | *OSM* | ENSG00000099985 | 22 | 30658819 | 30662829 |
| CXCL1 | 9,14E-12 | P09341 | *CXCL1* | ENSG00000163739 | 4 | 74735109 | 74737019 |
| CCL4 | 1,06E-13 | P13236 | *CCL4* | ENSG00000275302 | 17 | 34431220 | 34433014 |
| CD6 | 1,58E-03 | Q8WWJ7 | *CD6* | ENSG00000013725 | 11 | 60739113 | 60787848 |
| IL-18 | 2,16E-01 | Q14116 | *IL18* | ENSG00000150782 | 11 | 112013974 | 112034840 |
| TGF-alpha | 1,37E-04 | P01135 | *TGFA* | ENSG00000163235 | 2 | 70674412 | 70781147 |
| MCP-4 | 1,27E-03 | Q99616 | *CCR3* | ENSG00000181374 | 3 | 46283872 | 46308197 |
| CCL11 | 1,18E-06 | P51671 | *CCL11* | ENSG00000172156 | 17 | 32612687 | 32615199 |
| TNFSF14 | 1,40E-02 | O43557 | *TNFSF14* | ENSG00000125735 | 19 | 6663148 | 6670599 |
| MMP-1 | 5,25E-05 | P03956 | *MMP1* | ENSG00000196611 | 11 | 102660641 | 102668966 |
| IL-18R1 | 4,80E-05 | Q13478 | *IL18R1* | ENSG00000115604 | 2 | 102979093 | 103015230 |
| PD-L1 | 3,33E-01 | Q9NZQ7 | *CD274* | ENSG00000120217 | 9 | 5450503 | 5470567 |
| CXCL5 | 1,71E-07 | P42830 | *CXCL5* | ENSG00000163735 | 4 | 74861359 | 74864446 |
| HGF | 6,07E-01 | P14210 | *HGF* | ENSG00000019991 | 7 | 81331444 | 81399452 |
| IL-12B | 5,46E-04 | P29460 | *IL12B* | ENSG00000113302 | 5 | 158741791 | 158757481 |
| MMP-10 | 3,75E-02 | P09238 | *TIMP2* | ENSG00000166670 | 17 | 76849059 | 76921472 |
| IL-10 | 5,39E-01 | P22301 | *IL10* | ENSG00000136634 | 1 | 206940947 | 206945839 |
| TNF | 9,49E-03 | P01375 | *TNF* | ENSG00000228321 | 6 | 31544292 | 31546112 |
| CCL23 | 5,52E-01 | P55773 | *CCL23* | ENSG00000274736 | 17 | 34340097 | 34345005 |
| CD5 | 4,41E-04 | P06127 | *CD5* | ENSG00000110448 | 11 | 60869930 | 60895323 |
| CCL3 | 3,25E-16 | P10147 | *CCL3* | ENSG00000277632 | 17 | 34415603 | 34417506 |
| Flt3L | 6,92E-02 | P49771 | *FLT3LG* | ENSG00000090554 | 19 | 49977818 | 49990894 |
| CXCL10 | 2,21E-06 | P02778 | *CXCL10* | ENSG00000169245 | 4 | 76942269 | 76944689 |
| EN-RAGE | 3,67E-08 | P80511 | *S100A12* | ENSG00000163221 | 1 | 153346184 | 153348075 |
| CD40 | 1,04E-05 | P25942 | *CD40* | ENSG00000101017 | 20 | 44746906 | 44758384 |
| IFN-gamma | 2,05E-02 | P01579 | *INFG* | ENSG00000111537 | 12 | 68548548 | 68553527 |
| LIF | 2,02E-03 | P15018 | *LIF* | ENSG00000128342 | 22 | 30636436 | 30642840 |
| MCP-2 | 1,62E-14 | P80075 | *CCL8* | ENSG00000108700 | 17 | 32646066 | 32648421 |
| TNFRSF9 | 3,75E-01 | Q07011 | *TNFRSF9* | ENSG00000049249 | 1 | 7975931 | 8003225 |
| TWEAK | 1,41E-01 | O43508 | *TNFRSF12A* | ENSG00000006327 | 16 | 3070313 | 3072383 |
| CCL20 | 9,01E-04 | P78556 | *CCL20* | ENSG00000115009 | 2 | 228678558 | 228682280 |
| TNFB | 1,84E-01 | P01374 | *LTA* | ENSG00000226979 | 6 | 31540060 | 31542101 |
| CSF-1 | 9,98E-01 | P09603 | *CSF1* | ENSG00000184371 | 1 | 110453233 | 110473616 |

**Table S8.** The genome-wide significant pQTLs identified in the joint analysis show at least two associations with distinct proteins at P < 0.05/(10*32) = 1.56E-04, indicating potential pleiotropy between genome-wide significant pQTLs and measured inflammatory proteins. The cutoff reflects a conservative approach to multiple testing for all identified genome-wide significant pQTLs (10) with all tested proteins (32). Only associations at P < 1.56E-04 with the same allelic direction between 500FG and Lifelines DEEP cohort are shown in the table. SNP rs2488633 showed a single association at P thresdhold of 1.56E-04 and it was therefore excluded from the analysis of pleiotropic effects of genome-wide significant pQTLs on proteins. Chr: chromosome; HetPVal: P value of heterogeneity between 500FG and Lifelines DEEP cohort; Direction column shows the allelic direction between 500FG and Lifelines DEEP cohort.

| Protein | Chr | SNP | Allele1 | Allele2 | Z score | P-value | Direction | HetPVal |
|---|---|---|---|---|---|---|---|---|
| CCL4 | 1 | rs2501301 | t | c | 5,877 | 4,17E-09 | ++ | 0,3763 |
| MIP1alpha | 1 | rs2501301 | t | c | 4,284 | 1,84E-05 | ++ | 0,8404 |
| VEGFA | 3 | rs358011 | t | c | 5,84 | 5,23E-09 | ++ | 0,3636 |
| TGFalpha | 3 | rs358011 | t | c | 4,536 | 5,73E-06 | ++ | 0,6925 |
| OSM | 3 | rs358011 | t | c | 4,457 | 8,30E-06 | ++ | 0,5756 |
| CCL23 | 3 | rs358011 | t | c | 4,079 | 4,52E-05 | ++ | 0,305 |
| IL6 | 3 | rs358011 | t | c | 4,015 | 5,94E-05 | ++ | 0,873 |
| TNF | 3 | rs358011 | t | c | 3,928 | 8,56E-05 | ++ | 0,9293 |
| MCP1 | 3 | rs1398749 | a | g | -5,56 | 2,70E-08 | -- | 0,7 |
| IL8 | 3 | rs1398749 | a | g | -4,662 | 3,13E-06 | -- | 0,4147 |
| IL8 | 5 | rs392422 | a | g | -5,836 | 5,34E-09 | -- | 0,08817 |
| MCP1 | 5 | rs392422 | a | g | -3,847 | 1,20E-04 | -- | 0,1841 |
| IL8 | 8 | rs1699130 | c | g | -5,647 | 1,64E-08 | -- | 0,2991 |
| MCP1 | 8 | rs1699130 | c | g | -3,844 | 1,21E-04 | -- | 0,2684 |
| MCP2 | 10 | rs1572285 | t | c | -5,825 | 5,70E-09 | -- | 0,6382 |
| MCP3 | 10 | rs1572285 | t | c | -3,86 | 1,14E-04 | -- | 0,9336 |
| MCP2 | 10 | rs2488633 | a | g | -5,618 | 1,93E-08 | -- | 0,2206 |
| MCP3 | 16 | rs247426 | t | c | 5,522 | 3,36E-08 | ++ | 0,522 |
| IL6 | 16 | rs247426 | t | c | 4,448 | 8,68E-06 | ++ | 0,1817 |
| IL12B | 16 | rs247426 | t | c | 4,442 | 8,91E-06 | ++ | 0,8081 |
| MCP2 | 16 | rs247426 | t | c | 4,175 | 2,98E-05 | ++ | 0,1633 |
| CXCL1 | 16 | rs247426 | t | c | 3,94 | 8,13E-05 | ++ | 0,1007 |
| IL1alpha | 16 | rs247426 | t | c | 3,787 | 1,52E-04 | ++ | 0,5071 |
| CXCL9 | 3 | rs6771739 | t | c | 5,505 | 3,69E-08 | ++ | 0,09539 |
| IL1alpha | 3 | rs6771739 | t | c | 5,359 | 8,39E-08 | ++ | 0,9907 |
| IL18 | 3 | rs6771739 | t | c | 5,213 | 1,86E-07 | ++ | 0,1958 |
| TGFalpha | 3 | rs6771739 | t | c | 4,8 | 1,58E-06 | ++ | 0,7381 |
| IL6 | 3 | rs6771739 | t | c | 4,765 | 1,89E-06 | ++ | 0,5573 |
| CCL4 | 3 | rs6771739 | t | c | 4,641 | 3,47E-06 | ++ | 0,3647 |
| MMP1 | 3 | rs6771739 | t | c | 4,444 | 8,82E-06 | ++ | 0,9576 |
| MCP3 | 3 | rs6771739 | t | c | 4,42 | 9,87E-06 | ++ | 0,6512 |
| PDL1 | 3 | rs6771739 | t | c | 4,306 | 1,66E-05 | ++ | 0,3858 |
| CXCL10 | 3 | rs6771739 | t | c | 4,227 | 2,37E-05 | ++ | 0,6944 |
| OSM | 3 | rs6771739 | t | c | 4,044 | 5,26E-05 | ++ | 0,1792 |
| CCL23 | 3 | rs6771739 | t | c | 4,004 | 6,24E-05 | ++ | 0,9545 |
| Flt3L | 3 | rs6771739 | t | c | 3,996 | 6,43E-05 | ++ | 0,2451 |
| CXCL11 | 3 | rs6771739 | t | c | 3,986 | 6,71E-05 | ++ | 0,4731 |
| TNF | 3 | rs6771739 | t | c | 3,95 | 7,80E-05 | ++ | 0,4371 |
| CD40 | 3 | rs6771739 | t | c | 3,937 | 8,25E-05 | ++ | 0,1137 |
| VEGFA | 3 | rs6771739 | t | c | 3,935 | 8,30E-05 | ++ | 0,1148 |
| CXCL9 | 7 | rs34774255 | t | c | 5,515 | 3,50E-08 | ++ | 0,6533 |
| CXCL10 | 7 | rs34774255 | t | c | 5,034 | 4,80E-07 | ++ | 0,9942 |
| IL6 | 7 | rs34774255 | t | c | 4,724 | 2,31E-06 | ++ | 0,4156 |
| MCP3 | 7 | rs34774255 | t | c | 4,428 | 9,50E-06 | ++ | 0,3858 |
| CCL4 | 7 | rs34774255 | t | c | 4,255 | 2,09E-05 | ++ | 0,2908 |
| IL18 | 7 | rs34774255 | t | c | 4,173 | 3,01E-05 | ++ | 0,451 |

| Protein | Chr | SNP | Allele1 | Allele2 | Z score | P-value | Direction | HetPVal |
|---|---|---|---|---|---|---|---|---|
| IL1alpha | 7 | rs34774255 | t | c | 4,14 | 3,48E-05 | ++ | 0,4343 |
| TNF | 7 | rs34774255 | t | c | 4,107 | 4,02E-05 | ++ | 0,7667 |
| OSM | 7 | rs34774255 | t | c | 4,059 | 4,93E-05 | ++ | 0,1729 |
| IL12B | 7 | rs34774255 | t | c | 3,898 | 9,72E-05 | ++ | 0,6718 |
| CXCL1 | 7 | rs34774255 | t | c | 3,897 | 9,74E-05 | ++ | 0,529 |

**Table S9.** Seven SNPs or proxies (LD > 0.8), which are associated with infectious diseases, influence inflammatory proteins in response to *C. albicans*. Summary statistics of the SNPs associated with infectious diseases was extracted from the GWAS catalogue (https://www.ebi.ac.uk/gwas/). MAF calculated using a population consisted of 360 individuals from 500FG and 76 individuals from LifeLines DEEP cohort. chr: chromosome; MAF: minor allele frequency; OR: odds ratio.

| chr | SNP | Proxy SNP | Allele1 | Allele2 | MAF | P-value | Protein | Z score | Infectious disease | OR or beta | P value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | rs28561741 | rs4648739 | G | A | 0,27 | 7,49E-04 | VEGF-A | -3,371 | Severe influenza A (H1N1) infection | 16,85 | 8,00E-10 |
| 2 | rs76656639 | rs643684 | A | G | 0,19 | 8,33E-04 | CXCL5 | -3,342 | Severe influenza A (H1N1) infection | NA | 4,00E-08 |
| 3 | rs6787514 | rs6787514 | G | T | 0,23 | 4,41E-04 | CCL23 | 3,514 | Yeast infection | 0,0343 | 8,00E-06 |
| 5 | rs687444 | rs477687 | C | T | 0,17 | 3,31E-04 | CCL23 | 3,59 | AIDS progression | 1,55 | 6,00E-06 |
| 8 | rs7820437 | rs6992798 | A | G | 0,15 | 5,10E-04 | MIP-1alpha | 3,475 | Hepatitis B | 0,2364 | 9,00E-06 |
| 12 | rs7134436 | rs4766152 | T | G | 0,45 | 3,11E-04 | CXCL5 | 3,606 | Cytomegalovirus antibody response | 0,223 | 5,00E-06 |
| 15 | rs2881464 | rs2011905 | C | T | 0,21 | 2,64E-04 | MCP-1 | 3,649 | Cytomegalovirus antibody response | 0,222 | 2,00E-06 |

# CHAPTER 5

# The MHC locus and genetic susceptibility to autoimmune and infectious diseases

Vasiliki Matzaraki, Vinod Kumar,
Cisca Wijmenga and Alexandra Zhernakova

Matzaraki et al. Genome Biology (2017) 18:76
DOI 10.1186/s13059-017-1207-1

Genome Biology

REVIEW Open Access

# The MHC locus and genetic susceptibility to autoimmune and infectious diseases

Vasiliki Matzaraki[1], Vinod Kumar[1], Cisca Wijmenga[1,2*] and Alexandra Zhernakova[1]

**Abstract**

In the past 50 years, variants in the major histocompatibility complex (MHC) locus, also known as the human leukocyte antigen (HLA), have been reported as major risk factors for complex diseases. Recent advances, including large genetic screens, imputation, and analyses of non-additive and epistatic effects, have contributed to a better understanding of the shared and specific roles of MHC variants in different diseases. We review these advances and discuss the relationships between MHC variants involved in autoimmune and infectious diseases. Further work in this area will help to distinguish between alternative hypotheses for the role of pathogens in autoimmune disease development.

## Introduction

The major histocompatibility complex (MHC) locus, also known as the human leukocyte antigen (HLA) locus, spans around 4 Mbp on the short arm of chromosome 6 (6p21.3; Box 1). Molecules encoded by this region are involved in antigen presentation, inflammation regulation, the complement system, and the innate and adaptive immune responses, indicating the MHC's importance in immune-mediated, autoimmune, and infectious diseases [1]. Over the past 50 years, polymorphisms in the MHC locus have been shown to influence many critical biological traits and individuals' susceptibility to complex, autoimmune, and infectious diseases (Boxes 2 and 3). In addition to autoimmune and inflammatory diseases, the MHC has recently been found to play a role in some neurological disorders [2–6], implicating autoimmune components in these diseases.

* Correspondence: c.wijmenga@umcg.nl
[1]Department of Genetics, University of Groningen, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands
[2]Department of Immunology, KG Jebsen Coeliac Disease Research Centre, University of Oslo, PO Box 4950 Nydalen, 0424 Oslo, Norway

The genetic structure of the MHC is characterized by high levels of linkage disequilibrium (LD) compared to the rest of the genome, which means there are technical challenges in identifying MHC single nucleotide polymorphisms (SNPs), alleles, and amino acids. However, the recent availability of dense genotyping platforms, such as the custom-made Illumina Infinium SNP chip (Immunochip) [7], and of MHC reference panels has helped to fine-map the locus, improving our understanding of its disease associations and our ability to identify functional variants.

In this review, we discuss recent advances in mapping susceptibility variants in the MHC, using autoimmune and infectious diseases as examples (Boxes 2 and 3). We also discuss the relationships between the MHC variants involved in both autoimmune and infectious diseases and offer insights into the MHC-associated immune responses underlying disease onset and pathogenesis. Finally, we discuss future directions for studying genetic variation in the MHC and how learning about the variation at this locus will aid understanding of disease pathogenesis.

## Advances in mapping susceptibility variants in the MHC locus

Several computational and empirical challenges complicate the mapping of MHC susceptibility variants. One fundamental challenge is that the MHC has many sequence and structural variations [8], which differ between populations and complicate haplotype inference. Another is that high and extensive LD in the locus makes it difficult to identify causal and independent loci. Non-additive allelic effects in the MHC, and epistatic effects between the MHC and other loci, can also complicate inference of the underlying haplotype structure and disease susceptibility variants.

In recent years, large volumes of sequencing data have made it possible to impute MHC variation on a wide scale, thereby improving our understanding of variability at this locus and of the haplotype structures and

**Box 1. The major histocompatibility complex locus**

The major histocompatibility complex (MHC) was discovered in the mouse in 1936 [128]. It covers 0.13% of the human genome [1] and shows a high degree of polymorphism and extensive patterns of linkage disequilibrium (LD), which differ among populations. The large number of MHC alleles means each individual has a nearly unique set of peptide-presenting allotypic MHC molecules, and each MHC allotype confers the ability to bind different peptides. The MHC genes are classified into five subregions from the telomeric to the centromeric end: the extended class I, class I, class III, class II, and the extended class II regions [1]. The extended MHC region contains more than 400 annotated genes and pseudogenes that extend beyond the boundaries defining the MHC.

The class I region includes the three classic human leukocyte antigen (HLA) gene loci: HLA-A, HLA-B, and HLA-C; three non-classic HLA-E, HLA-F, and HLA-G gene loci, which show limited polymorphism compared to the classic class I loci; and other related non-coding genes and pseudogenes [1]. The main function of HLA class I molecules, which are expressed in all nucleated cells, is to present non-self antigens derived from intracellular sources, such as viruses, to $CD8^+$ T cells (cytotoxic T cells), which then kill the antigen-presenting cells (APCs) [129]. $CD8^+$ T cells interact with the cognate peptide-MHC I complexes via their T-cell receptor (TCR) and co-receptor molecule CD8.

The class II region includes the classic gene loci HLA-DP, HLA-DQ, and HLA-DR and also the non-classic HLA-DO and HLA-DM loci [1]. The classic genes are expressed on the surface of professional APCs, which take up antigens derived from extracellular sources [130], such as bacteria or food, and present them to $CD4^+$ T helper cells. This leads to the secretion of various small proteins, including cytokines, which regulate other immune cells such as macrophages or B cells. In turn, macrophages can destroy ingested microbes, and activated B cells can secrete antibodies. $CD4^+$ T cells interact with the cognate peptide-MHC II complexes via their TCR and the co-receptor molecule CD4. Non-classic molecules are exposed in internal membranes in lysosomes, which help load antigenic peptides on to classic MHC class II molecules.

The class III region contains genes involved in inflammation, for example, complement cascades (C2, C4, CFB), and in cytokine production (TNF, LTA, LTB), as well as many other genes of non-immune or unknown function that may not be involved in inflammation [1].

Overall, classic MHC I and II molecules present peptides for T-cell surveillance and are, therefore, critical for priming the cellular adaptive immune responses.

enabling reference panels to be created. The availability of accurate reference panels and a large number of genotyped individuals has allowed the identification of independent variants and improved our understanding of their contribution to disease heritability and pathways underlying disease biology [9, 10].

### Advances in laboratory-based mapping of MHC variation

Increased throughput, accuracy, and read length in next-generation sequencing (NGS) technologies, as well as the development of user-friendly bioinformatics tools, have enabled higher resolution MHC typing [11]. For instance, whole-genome sequencing (WGS) was successfully used to type HLA-A alleles at full resolution in 1070 healthy Japanese individuals [12] and to fully evaluate HLA-E variability in West African populations [13]. However, the main problem with MHC sequencing using current technologies is the relatively short read lengths, which limit the amount of allelic data that can be generated at a high resolution. Long-range PCR amplification approaches, such as the use of PacBio systems for single molecule real-time sequencing, significantly increase read-length and the resolution for typing MHC alleles [14]. In a comparison of MHC typing in an Indian population using sequence-specific primers, NGS (Roche/454) and single molecule sequencing (PacBio RS II) platforms, higher resolution typing was achieved for MHC class I (HLA-A, HLA-B, and HLA-C ) and class II genes (HLA-DRB1 and HLA-DQB1) using the PacBio platform, with a median read length of 2780 nucleotides [15].

High-density SNP panels, such as the Immunochip platform [7], which has been widely implemented in immunogenetics studies, are a cheaper, faster, and easier alternative to genotyping than direct MHC typing and NGS methods. The Immunochip contains a dense panel of SNPs from the MHC locus, which enables missing classic MHC variants to be inferred in silico, where the imputation is based on the haplotype structure present in large reference panels (Fig. 1). This fine-mapping approach has been used for several autoimmune and inflammatory diseases (Table 1) and for a few infectious diseases (Additional file 1), thereby allowing comprehensive interrogation of the MHC. Moreover, population-specific reference panels made by deep sequencing and used to impute genotypes allow identification of very rare variants and novel single-nucleotide variants in the human genome. This is illustrated by a recent study in which the authors first built a Han Chinese MHC-specific database by deep sequencing the region in 9946 patients with psoriasis and 10,689 healthy controls, and then used this reference panel to impute genotype data to fine-map psoriasis-associated variants [16]. Notably, functional variants in non-coding regions can be

**Box 2. Clinical characteristics and prevalence of autoimmune diseases in Europeans**

**Rheumatoid arthritis**

Chronic inflammation of synovial joints, with a prevalence of 0.5–1%. In some individuals, rheumatoid arthritis can damage a wide variety of body systems, including the skin, eyes, lungs, heart, and blood vessels.

**Celiac disease**

Chronic inflammation of the intestine triggered by gluten peptides in the diet and leading to flattening of the mucosa. Prevalence is 0.5–2%.

**Psoriasis**

An inflammatory skin condition characterized by rapid growth and skin cell reproduction. The disease trigger is unknown. The prevalence is 0.5–1% worldwide, but it is higher (2%) in Europeans.

**Ankylosing spondylitis**

A chronic, degenerative, and inflammatory form of arthritis, primarily affecting the spine and sacroiliac joints, and eventually leading to spinal fusion. This makes the spine less flexible and can result in a hunchback posture. It has a prevalence of 0.025%.

**Systemic lupus erythematosus**

Chronic inflammation that can affect any part of the body, although it often attacks the heart, joints, skin, lungs, blood vessels, liver, kidneys, and nervous system. It has a prevalence of 0.04–0.12%.

**Type 1 diabetes**

Characterized by the destruction of pancreatic beta-cells, leading to insufficient release of insulin from the pancreas. It has a prevalence of 0.2–0.3%.

**Multiple sclerosis**

Characterized by autoimmune attack on the central nervous system, leading to demyelination of neurons, and potentially debilitating physical and mental symptoms. It has a prevalence of 0.02%.

**Graves' disease**

An autoimmune thyroid disorder leading to the overproduction of thyroid hormones (hyperthyroidism). Graves' disease occurs in about 0.5% of males and 3% of females [131]. It is the most common cause of hyperthyroidism in the USA, affecting about 1 in 200 people (0.5%) according to the National Institutes of Health (https://ghr.nlm.nih.gov/condition/graves-disease#statistics).

**Inflammatory bowel disease**

A group of intestinal disorders involving chronic inflammation of the digestive tract. The two most common types of inflammatory bowel disease are Crohn's disease (CD), which is characterized by inflammation of any part of the digestive tract, and ulcerative colitis (UC), in which the inflammation is mostly localized in the large intestine. In Europe, CD has a prevalence varying from 0.00015 to 0.2%, whereas the prevalence of UC varies from 0.0024 to 0.3% [132].

**Dermatomyositis**

A rare idiopathic myopathy characterized by inflammation, primarily of the muscles and skin. It may also affect the joints, esophagus, lungs, and heart. The disease incidence ranges from 1.2 to 17 new cases per 1,000,000 inhabitants, with a prevalence between 0.005 and 0.011% [133].

identified, as shown in a Japanese cohort of 1070 healthy individuals [12]. These variants would be impossible to discover using SNP microarrays or low coverage sequencing on the same sample size (Fig. 1, Table 1).

MHC associations revealed by genome-wide association studies (GWAS) can often not be fine-mapped to a single allele at a single locus; rather they comprise independent effects from multiple loci (see "Role of MHC variants in human diseases"). The presence of these multiple, independent effects highlights the heterogeneous nature within and between diseases, which may lead to varying immunological responses. Fine-mapping has also shown that autoimmune diseases share MHC alleles and hence molecular pathways, which are likely to represent targets for shared therapies. For instance, the major associations within MHC class II across autoimmune diseases imply that modulating T-cell receptor (TCR) activation by using peptide-bearing MHC molecules on antigen-presenting cells (APCs) could be therapeutically useful [17]. Shared MHC genetic factors have also been observed between autoimmune and infectious diseases, suggesting that human genetic architecture has evolved in response to natural selection as determined by various infectious pathogens [18].

## Advances in computational approaches for mapping MHC variation

Long-range LD between loci and SNP markers across the MHC offers an alternative approach to interrogate functional MHC variation through imputation. The development of different imputation tools using population-specific reference panels has enhanced the interpretation of genotype data derived from genome-wide platforms. MHC imputation is done using reference panels containing both genetic information and classic HLA serotyping, thus allowing identification of MHC allelic and amino acid variants. It is advantageous to impute allele and amino acid variants in the MHC because background sequence diversity causes the binary SNP concept to fail, technically speaking, while

**Box 3. Infectious diseases and infection-related GWAS phenotypes**

**HIV infection**

Infection by the lentiviral human immunodeficiency virus (HIV; a subgroup of retrovirus) is a global public health issue. According to the World Health Organization (WHO), 36.7 million people were living with HIV at the end of 2015 (http://www.who.int/). The virus attacks human immune cells. Over time, HIV infection develops into acquired immunodeficiency syndrome (AIDS), a condition characterized in humans by progressive loss of immune function and leading to life-threatening opportunistic infections and cancers.

**Dengue shock syndrome**

Dengue shock syndrome is the most dangerous and severe complication of infection with the dengue virus. It is characterized by increased vascular permeability, together with myocardial dysfunction and dehydration. Dengue virus is a single, positive-stranded RNA virus of the *Flaviviridae* family; it is mainly transmitted by mosquitos. Dengue is found in tropical and sub-tropical climates, mostly in urban and semi-urban regions. The WHO estimates that about half of the world's population is now at risk.

**Hepatitis B virus infection**

Hepatitis B virus (HBV) is a double-stranded DNA virus belonging to the *Heoadnaviridae* family. The virus can cause both acute and chronic infections. Chronic infection with HBV leads to serious liver disease, often progressing to liver cirrhosis and hepatocellular carcinoma [134]. The WHO estimates that about 240 million people live with chronic HBV infection worldwide, with the highest prevalence seen in Africa and Asia.

**Hepatitis C virus infection**

Hepatitis C virus (HCV) is a single-stranded RNA virus of the *Flaviviridae* family. It has the same target as HBV—the liver—and can cause both acute and chronic infections. The WHO estimates that 130–150 million people worldwide have chronic infection; many of these will go on to develop liver cirrhosis or liver cancer. The prevalence of HCV infections is highest in Africa and Central and East Asia.

**Human papillomavirus virus infection**

Human papillomavirus virus (HPV) covers a highly diverse group of DNA papillomaviruses that are common worldwide. They can infect either mucosal or cutaneous epithelia but, in most cases, can be cleared by the human immune system. If infection persists, certain high-risk mucosal types (e.g., HPV16 and HPV18) can lead to cervical cancer and other anogenital and oropharyngeal cancers.

**Leprosy**

Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*; it mainly affects the skin, peripheral nerves, mucosa of the upper respiratory tract, and eyes. It is curable using multidrug therapy, which the WHO has made available free of charge to patients worldwide since 1995. The WHO reports on leprosy in 121 countries and territories, but not in Europe, and gave a prevalence of 175,554 cases at the end of 2014. Thus, leprosy remains a serious public health problem, especially in developing countries. Leprosy is classified into five distinct clinical subtypes. At one end of the spectrum, tuberculoid leprosy (TT) is characterized by fewer lesions and resistance to mycobacteria proliferation, caused by a robust Th1 antigen-specific cellular response. In contrast, at the other end of the spectrum, lepromatous leprosy (LL) is characterized by numerous lesions and proliferation of mycobacteria because of a weak or absent cellular immune response and a dominant Th2 response. Between TT and LL, there is a range of intermediate forms and manifestations [135].

**Tuberculosis**

Tuberculosis is caused by *Mycobacterium tuberculosis*; it most often affects the lungs but can affect other parts of the body. It is one of the top infectious killers worldwide, with over 95% of cases and deaths occurring in developing countries. Sub-Saharan Africa has the highest disease prevalence. Pulmonary tuberculosis is the only transmissible form of the disease and the most common form seen in all ages (WHO, Global Tuberculosis Report 2016; http://apps.who.int/iris/bitstream/10665/250441/1/9789241565394-eng.pdf?ua=1). In 2015, the WHO estimated 10.4 million new (incident) tuberculosis cases worldwide.

**Leishmaniasis**

Leishmaniasis is caused by the protozoan *Leishmania* parasites, which are transmitted to humans by infected female sandfly bites. The disease has three forms: visceral (also known as kala-azar, the most serious form of the disease), cutaneous (the most common), and mucocutaneous. It is classified as a neglected tropical disease, and the WHO estimates that there are 900,000 to 1.3 million new cases and 20,000 to 30,000 deaths annually. Visceral leishmaniasis is endemic in the Indian subcontinent and in East Africa, while cutaneous leishmaniasis is most common in the Americas, the Mediterranean basin, the Middle East, and Central Asia. Cases of mucocutaneous leishmaniasis occur in South America (Bolivia, Peru, and Brazil; WHO Fact Sheet, updated September 2016; (http://www.who.int/mediacentre/factsheets/fs375/en/).

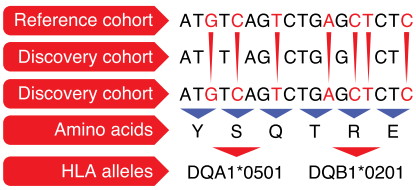



**Fig. 1** Major histocompatibility complex imputation. A reference cohort of subjects for whom both genetic information and classic human leukocyte antigen (HLA) typing is available can be used to infer the missing (untyped) genotypes and amino acids in a discovery cohort. This allows imputed variants to be tested for their associations with a disease of interest. The figure shows imputation points to classic alleles associated with celiac disease risk in the MHC region on chromosome 6. *Y* tyrosine, *S* serine, *Q* glutamine, *T* threonine, *R* arginine, *E* glutamic acid

many SNPs have more than two alleles and various amino acids can be contained in the same position. For instance, six possible amino acid variants at position 11 in the HLA-DRB1 gene show the strongest association to rheumatoid arthritis (RA) [19]. Two of these (valine and leucine) confer susceptibility to RA, whereas the other four (asparagine, proline, glycine and serine) are protective.

Several tools allowing imputation of classic HLA alleles at four-digit resolution are now available for MHC imputation analysis; the most common are SNP2HLA [20], HLA*IMP:01 [21], and an improved HLA*IMP:02 [22]. HLA*IMP:02 outperforms HLA*IMP:01 on heterogeneous European populations and it increases the power and accuracy in cross-European GWAS [22]. Missing data are also better tolerated in HLA*IMP:02, while SNP genotyping platforms must be selected in HLA*IMP:01 [21, 22]. SNP2HLA not only imputes classic alleles but also amino acids by using two European reference panels, one based on data from HapMap-CEPH (90 individuals), and the other on the Type 1 Diabetes Genetics Consortium (T1DGC) study [20]. Another tool, HLA-VBSeq, allows imputation of MHC alleles at full resolution from whole-genome sequence data [23]. HLA-VBSeq does not require prior knowledge of MHC allele frequencies and can therefore be used for samples from genetically diverse populations [23]. It has successfully typed HLA-A alleles at full resolution in a Japanese population and identified rare causal variants implicated in complex human diseases [12].

One commonly used European reference panel for imputation is the T1DGC panel, which covers SNP genotyping and classic HLA serotyping information for 5225 unrelated individuals [20]. Similar population-specific reference panels have been developed for non-European studies to investigate the risk of psoriasis in Chinese populations [16] and of Graves' disease and RA in Japanese populations. The panels have also been used to impute MHC alleles and amino acids for East Asian and Korean populations [24–26].

Using a single reference genome for regions like the MHC, which has substantial sequence and structural diversity, results in poor characterization. To counteract this, an algorithm was developed to infer much of the variation in the MHC; it allows genome inference from high-throughput sequencing data using known variation represented in a population reference graph (PRG) [27]. Specifically, the PRG constructed for the MHC combined eight assembled haplotypes, the sequences of known classic HLA alleles, and 87,640 SNP variants from the 1000 Genomes Project [28]. This approach is considered to be an intermediate step between de novo assembly and mapping to a single reference, but requires careful attention to the variation included in the PRG [27].

Despite the development of new tools to investigate MHC variation, the robustness of imputation depends largely on the reference panel and SNP selection. The frequency of alleles can differ between populations, thus highlighting the need to use population-specific reference panels to impute MHC alleles and amino acids. Additionally, the use of many samples is possible for analyzing the non-additive effects of MHC alleles on a wide scale, as described by Lenz et al. for celiac disease (CeD), psoriasis, and type 1 diabetes (T1D) [29]. These non-additive effects could explain our inability to identify susceptibility variants. However, one important limitation of existing imputation methods is that they are limited to the classic MHC alleles and their amino acids. Another limitation is that accuracy is lower for low frequency or rare variants [20, 30]; this can be improved by increasing the reference panel size, together with the use of deep sequencing data. Ascertainment bias and lower LD also make it challenging to impute MHC variants in some non-European populations, such as Africans.

MHC genetic variation mediates susceptibility to a wide range of complex diseases, including infectious and autoimmune diseases. The large volume of data generated by recent GWAS has provided an excellent opportunity to apply imputation tools used to fine-map MHC associations to classic alleles and amino acids, as described below for autoimmune diseases. Overall, MHC imputation has proved to be a robust and cost-effective way to identify causal genes underlying disease pathogenesis. Ultimately, knowing the causal genes will help explain disease heritability and lead to a better understanding of the molecular pathways involved in disease pathogenesis. Such work helps to pinpoint potential therapeutic targets.

**Table 1** Major histocompatibility complex (MHC) associations to autoimmune diseases, as described by fine-mapping studies

| Disease | Classic associated locus | Fine-mapping: top locus | HLA class for top association | Fine-mapping: independent associations at $P < 5 \times 10^{-8}$ | HLA class for independent association | Population | Sample size | Reference |
|---|---|---|---|---|---|---|---|---|
| RA | Shared epitope (HLA-DRB1 AA70-74) | HLA-DRB1: AA 11, 13, 71, 74 | II | HLA-B AA 9 | I | European | 5018 individuals with seropositive RA, 14,974 unaffected controls | [19] |
| | | | | HLA-DPB1 AA 9 | II | | | |
| | | HLA-DRB1: AA 11, 57, 74 | II | | | Chinese, Korean | 2782 seropositive RA cases, 4315 controls | [34] |
| | | HLA-DRB1: AA 11, 71, 74 | II | HLA-DPB1 AA 84 | II | Japanese | 6244 cases, 23,731 controls | [25] |
| | | | | HLA-DOA (rs378352) | II | | | |
| | | | | HLA-B 40:02 | I | | | |
| CeD | HLA DQ locus (DQ2, DQ8) | HLA-DQA1: AA 25 and 47; HLA-DQB1: AA 57 and 74 | II | AA9 in HLA-DPβ1 | II | European | 12,016 cases, 11,920 controls | [35] |
| | | | | rs2301226 (near HLA-DPB1 gene) | II | | | |
| | | | | HLA-B*08:01 | I | | | |
| | | | | HLA-B*39:06 | I | | | |
| | | | | rs1611710 (HLA-F) | I | | | |
| Ps | HLA-Cw6 (HLA-C 0602) | HLA-C*06:02 | I | HLA-C*12:03 | I | European | 9247 patients, 13,589 controls | [36] |
| | | | | HLA-B amino acid 67 | I | | | |
| | | | | HLA-B amino acid 9 | I | | | |
| | | | | HLA-A amino acid position 95 | I | | | |
| | | | | HLA-DQα1 amino acid position 53 | II | | | |
| AS | HLA-B27 | HLA-B*27:02 and B*27:05 | I | HLA-A*02:01 | I | European | 9069 AS cases, 13,578 controls | [37] |
| | | | | HLA-DPB1 (rs1126513) | II | | | |
| | | | | HLA-DRB1*01:03 | II | | | |
| | | | | other HLA-B alleles (B*07:02) (protective), B*13:02, B*40:01, B*40:02, B*47:01, B*51:01 and B*57:01 (protective) | I | | | |
| SLE | HLA-DRB1 (*03:01 and *15:01); TNF-308G/A; C4A null allele | HLA-DRB1 AA 11, 12, 26 | II | HLA-DQB1*02 | II | Korean | 849 SLE cases, 4493 controls | [38] |
| | | HLA-DRB1*15:01-HLA-DQB1*06:02 | II | | | | | |
| | | DRB1*0301 | II | rs419788 in intron 6 of the SKIV2L | III | European | 314 trios | [39] |
| | | rs1150753 in intronic region of TNXB | III | HLA-DRB1*03:01 | II | European | 3701 cases, 12,110 controls | [40] |
| | | | | HLA-DRB1*08:01 | II | | | |
| | | | | HLA-DQA1*01:02 | II | | | |
| | | | | rs8192591 (upstream of NOTCH4) | III | | | |
| | | | | rs2246618 near MICB | I | | | |

**Table 1** Major histocompatibility complex (MHC) associations to autoimmune diseases, as described by fine-mapping studies *(Continued)*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | HLA-DRB1 AA 13, 11; HLA-DRB1*15:01-HLA-DQB1*06:02 | II | | | 6 Asian cohorts (from China, Korea, Japan) | 4478 cases, 12,656 controls | [41] |
| T1D | DRB1, DQB1, in particular heterozygosity for HLA-DR3/HLA-DR4 | HLA-DQβ1 position 57; HLA-DRβ1 positions 13; HLA-DRβ1 positions 71 | II | HLA-B*39:06<br>HLA-B*18:01<br>HLA-B*50:01<br>HLA-DPB1*04:02<br>HLA-DPB1*01:01<br>HLA-A AA62<br>HLA-A*03<br>HLA-A*24:02 | I<br>I<br>I<br>II<br>II<br>I<br>I<br>I | European | 8095 cases, 10,737 controls | [43] |
| | | DRB1, DQB1 | II | HLA-B*39<br>HLA-B*13<br>HLA-B*50<br>HLA-B*18<br>HLA-B*38 (protective)<br>HLA-A*24 | I<br>I<br>I<br>I<br>I<br>I | European | 2300 families | [42] |
| MS | HLA-DRB1*15:01 | HLA-DRB1*15:01 | II | DRB1 alleles: *03:01, *13:03, *04:04, *04:01, and *14:01 | II | European | 5091 cases, 9595 controls | [45] |
| | | | | HLA-A*02:01 | I | | | |
| | | | | AA 65 of HLA-DPβ1 | II | | | |
| | | | | rs2516489 in MICB-LST1 locus | III | | | |
| | | | | HLA-B*37 | I | | | |
| | | HLA-DRB1*15:01 | II | HLA-DRB1*03:01, *13:03, *08:01 | II | European | 17,465 cases, 30,385 controls | [44] |
| | | | | HLA-DQB1*03:02 | II | | | |
| | | | | HLA-A*02:01 (protective) | I | | | |
| | | | | HLA-B*44:02, *38:01, *55:01 (protective) | I | | | |
| CD | HLA class II and HLA-C | HLA-DRB1*01:03 | II | HLA-C*06:02 | I | European | 18,405 cases, 34,241 controls | [46] |
| UC | HLA class II | rs6927022(HLA-DQA1); HLA-DRB1*01:03 | II | HLA-C*12:02 | I | European | 14,308 cases, 34,241 controls | |
| Graves' disease | DRB1*03:01-DQA1*05:01-DQB1*02:01 (DR3-DQ2 haplotype), HLA-C | HLA-DPB1 AA35, AA55, DPB1*05:01 | II | HLA-DPβ1 AA9 | II | Japanese | 1956 cases, 7047 controls | [24] |
| | | | | HLA-A AA9 | I | | | |
| | | | | HLA-B AA45, AA67 | I | | | |
| DM | HLA-DPB1 | HLA-DPB1*17 | II | HLA-DBP1 (rs7750458-A) | II | Han Chinese | 127 cases, 1566 controls | [47] |

*RA* rheumatoid arthritis, *CeD* celiac disease, *Ps* psoriasis, *AS* ankylosing spondylitis, *SLE* systemic lupus erythematosus, *T1D* type 1 diabetes, *MS* multiple sclerosis, *CD* Crohn's disease, *UC* ulcerative colitis, *DM* dermatomyositis

## Role of MHC variants in human diseases

### Insights into MHC susceptibility for autoimmune diseases: fine-mapping results, epistasis, and disease biology

Associations between the MHC and autoimmune diseases reported in the 1970s were some of the earliest described genetic associations [31, 32], and they remain the strongest risk factors for autoimmune diseases. After the development of wide-screen genotyping platforms and imputation pipelines, MHC imputation and fine-mapping were performed in European and Asian populations for most common autoimmune diseases, including RA [19, 25, 33, 34], CeD [35], psoriasis [36], ankylosing spondylitis (AS) [37], systemic lupus erythematosus (SLE) [33, 38–41], T1D [42, 43], multiple sclerosis (MS) [44, 45], Graves' disease [24], inflammatory bowel disease (IBD) [46], and dermatomyositis (DM) [47]. Table 1 shows the main associated variants and independently associated loci for autoimmune diseases.

In 2012, a pioneering MHC fine-mapping study, performed in individuals of European ancestry with RA [19], confirmed the strongest association with the class II HLA-DRB1 gene, as well as other independent associations. Previously an increased risk of RA was reported for a set of consensus amino acid sequences at positions 70–74 in the HLA-DRB1 gene, known as the "shared epitope" locus [48]. The imputed data revealed the most significant associations were with two amino acids at position 11, located in a peptide-binding groove of the HLA-DR heterodimer. This suggested a functional role for this amino acid in binding the RA-triggering antigen. Similar fine-mapping studies followed for other autoimmune diseases (Table 1).

In general, in most autoimmune diseases, fine-mapping strategies have confirmed the main associated locus reported by serotype analysis within a certain MHC locus. Such strategies have also allowed identification of specific allelic variants or amino acids, as well as independent variants in different HLA classes. For instance, in CeD, the strongest association was with the known DQ-DR locus, and five other independent signals in classes I and II were also identified. CeD is the only autoimmune disease for which the antigen, gluten, is known and well studied. Gluten is a dietary product in wheat, barley, and rye. It is digested in the intestine and deamidated by tissue transglutaminase enzymes such that it perfectly fits the binding pockets of a particular CeD-risk DQ heterodimer (encoded by the DQ2.2, DQ2.5, and DQ8 haplotypes). This association was confirmed by MHC fine-mapping, which indicated roles for four amino acids in the DQ genes with the strongest independent associations to CeD risk [35]. Similarly, the main associations were determined for T1D, MS, and SLE within the MHC class II locus (the associations for these three diseases are to a particular HLA-DQ-DR haplotype), and there are also independent, but weaker associations with the class I and/or III regions. In DM, fine mapping in an Asian population identified MHC associations driven by variants located around the MHC class II region, with HLA-DP1*17 being the most significant [47]. In contrast, the primary and strongest associations in psoriasis and AS were to MHC class I molecules, while independent associations to the class I locus were also reported for IBD and Graves' disease. Class III variants are weakly implicated in autoimmune diseases, but several associations in the MHC class III region were seen for MS; for instance, the association to rs2516489 belonging to the long haplotype between *MICB* and *LST1* genes. The association signal to rs419788-T in the class III region gene *SKIV2L* has also been implicated in SLE susceptibility, representing a novel locus identified by fine-mapping in UK parent–child trios [39]. An independent association signal to class III was also identified (rs8192591) by a large meta-analysis of European SLE cases and controls and, specifically, upstream of *NOTCH4* [40]. However, further studies are needed to explain how these genetic variations contribute to predisposition to SLE.

In addition to identifying independent variants, MHC fine-mapping studies permit analysis of epistatic and non-additive effects in the locus. These phenomena occur when the effect of one allele on disease manifestation depends on the genotype of another allele in the locus (non-additive effect), or on the genotype of the "modifier" gene in another locus (epistasis). Non-additive MHC effects were established in CeD, in which knowing gluten was the causal antigen offered an advantage in investigating the antigen-specific structure of the DQ-heterodimer. CeD risk is mediated by the presence of several HLA-DQ haplotypes, including the DQ2.5, DQ2.2, and DQ8 haplotypes, which form the specific pocket that efficiently presents gluten to T cells. These haplotypes can be encoded either in *cis*, when both DQA1 and DQB1 are located on the same chromosome, or *in trans*, when they are located on different chromosomes. Some DQ allelic variants confer susceptibility to CeD only in combination with certain other haplotypes, forming a CeD-predisposing *trans*-combination. For example, HLA-DQA1*0505-DQB1*0301 (DQ7) confers risk to CeD only if it is combined with DQ2.2 or DQ2.5, contributing to the formation of susceptible haplotypes *in trans*. In particular, DQ7/DQ2.2 heterozygosity confers a higher risk for CeD than homozygosity for either of these alleles, and is an example of a non-additive effect for both alleles.

Unlike CeD, the exact haplotypes and their associated properties remain unknown for most other autoimmune diseases; therefore, analyzing non-additive effects might yield new insights into potentially disease-causing

antigens. Lenz et al. provided evidence of significant non-additive effects for autoimmune diseases, including CeD, RA, T1D, and psoriasis, which were explained by interactions between certain classic HLA alleles [29]. For instance, specific interactions that increase T1D disease risk were described between HLA-DRB1*03:01-DQB1*02:01/DRB1*04:01-DQB1*03:02 genotypes [49] and for several combinations of the common HLA-DRB1, HLA-DQA1, and HLA-DQB1 haplotypes [43]. In AS, epistatic interaction was observed for combinations of HLA-B60 and HLA-B27, indicating that individuals with the HLA-B27+/HLA-B60+ genotype have a high risk of developing AS [50]. Moreover, a recent study in MS found evidence for two interactions involving class II alleles: HLA-DQA1*01:01-HLA-DRB1*15:01 and HLA-DQB1*03:01-HLA-DQB1*03:02, although their contribution to the missing heritability in MS was minor [44].

Epistatic interactions between MHC and non-MHC alleles have also been reported in several autoimmune diseases, including SLE, MS, AS, and psoriasis. For instance, in a large European cohort of SLE patients, the most significant epistatic interaction was identified between the MHC region and cytotoxic T lymphocyte antigen 4 (*CTLA4*) [9], which is upregulated in T cells upon encountering APCs. This highlights that appropriate antigen presentation and T-cell activation are important in SLE pathogenesis [9]. Notably, interactions between MHC class I and specific killer immunoglobulin receptor (KIR) genes are important in predisposition to autoimmune diseases such as psoriatic arthritis, scleroderma, sarcoidosis, and T1D [51–54]. KIR genes are encoded by the leukocyte receptor complex on chromosome 19q13 and expressed on natural killer cells and subpopulations of T cells [55]. Finally, epistatic interactions between MHC class I and *ERAP1* have been described for AS, psoriasis, and Behçet's disease [10].

Association of novel MHC variants and identification of interaction effects within the MHC are increasing our understanding of the biology underlying autoimmune and inflammatory diseases. Fine-mapping the main associated locus within HLA-DQ-DR haplotypes has allowed determination of the key amino acid positions in the DQ or DR heterodimer. Pinpointing specific amino acids leads to a better understanding of the structure and nature of potential antigens for autoimmune or inflammatory diseases, and these can then be tested through binding assays and molecular modeling. The fact that these positions are located in peptide-binding grooves suggests they have a functional impact on antigenic peptide presentation to T cells, either during early thymic development or during peripheral immune responses [19]. In addition, analysis of non-additive effects in MHC-associated loci offers the possibility to identify antigen-specific binding pockets and key amino acid sequences. For example, identification of the protective, five-amino acid sequence DERAA as a key sequence in the RA-protective HLA-DRB1:13 allele, and its similarity to human and microbial peptides, led to identification of (citrullinated) vinculin and some pathogen sequences as novel RA antigens [56].

The identification of independent signals in MHC classes I and III for many autoimmune diseases implies that these diseases involve novel pathway mechanisms. For example, association of CeD to class I molecules suggests a role for innate-like intraepithelial leucocytes that are restricted to class I expression and that are important in epithelial integrity and pathogen recognition [57]. Class I associations to RA, T1D, and other autoimmune diseases suggest that $CD8^+$ cytotoxic cells are involved in disease pathogenesis, as well as $CD4^+$ helper T cells.

Discovering the epistatic effects of MHC and non-MHC loci can also shed light on disease mechanisms. For example, *ERAP1* loss-of-function variants reduce the risk of AS in individuals who are HLA-B27-positive and HLAB-40:01-positive, but not in carriers of other risk haplotypes [37]. Similar epistatic effects were also observed for psoriasis, such that individuals who carry variants in ERAP1 showed an increased risk only when they also carried an HLA-C risk allele [58]. In line with these observations, mouse studies have shown that *ERAP1* determines the cleavage of related epitopes in such a way that they can be presented by the HLA-B27 molecule [37]. Confirming that certain epitopes must be cleaved by ERAP1 to be efficiently presented by $CD4^+$ and $CD8^+$ cells will be a critical step in identifying specific triggers for autoimmune diseases.

The recent discoveries of genetic associations between MHC alleles and autoimmune diseases are remarkable and offer the potential to identify disease-causing antigens. This would be a major step towards developing new treatments and preventing disease. However, we still do not understand exactly how most associated alleles and haplotypes work, and extensive functional studies are needed to clarify their involvement in disease.

### Explained heritability by independent MHC loci for autoimmune diseases

Heritability is an estimation of how much variation in a disease or phenotype can be explained by genetic variants. Estimating heritability is important for predicting diseases but, for common diseases, it is challenging and depends on methodological preferences, disease prevalence, and gene–environment interactions that differ for each phenotype [59]. It is therefore difficult to compare

heritability estimates across diseases. Nevertheless, for many diseases, estimates have been made as to how much phenotypic variance can be explained by the main locus and by independent MHC loci [29].

For autoimmune diseases with a main association signal coming from a class II locus, the reported variance explained by MHC alleles varies from 2 – 30% [9]. The strongest effect is reported for T1D, in which the HLA-DR and HLA-DQ haplotypes explain 29.6% of phenotypic variance; independently associated loci in HLA-A, HLA-B, and HLA-DPB1 together explain about 4% of the total phenotypic variance, while all other non-MHC loci are responsible for 9% [60]. Similarly, in CeD, which has the same main associated haplotype as T1D, the HLA-DQ-DR locus explains 23 – 29% of disease variance (depending on the estimated prevalence of disease, which is 1 – 3%), whereas other MHC alleles explain 2 – 3%, and non-MHC loci explain 6.5 – 9% [35]. In seropositive RA, 9.7% of phenotypic variance is explained by all the associated DR haplotypes, whereas a model including three amino acid positions in DRB1, together with independently associated amino acids in HLA-B and HLA-DP loci, explains 12.7% of the phenotypic variance [19]. This indicates that non-DR variants explain a proportion of heritability comparable to that in other non-MHC loci (4.7 – 5.5% in Asians and Europeans) [19]. The non-additive effects of DQ-DR haplotypes can also explain a substantial proportion of phenotypic variance: 1.4% (RA), 4.0% (T1D), and 4.1% (CeD) [29]. In MS, the major associated allele, DRB1*15:01, accounts for 10% of the phenotypic variance, whereas all the alleles in DRB1 explain 11.6%. A model including all of the independent variants (and those located in classes I, II, and III) explains 14.2% of the total variance in MS susceptibility [45].

In SLE, the proportion of variance explained by the MHC is notably lower, at only 2% [41], and is mostly due to class II variants. In IBD, the association with MHC is weaker than in classic autoimmune diseases, with a lower contribution seen in Crohn's disease (CD) than in ulcerative colitis (UC) [61]. The main and secondary variants can now explain 3.1% of heritability in CD and 6.2% in UC, which is two to ten times greater than previously attributed by main effect analysis in either disease (0.3% in CD and 2.3% in UC for the main SNP effect) [46]. Among all the diseases discussed here, the main effect of the associated haplotype is far stronger than the independent effects from other loci (with the exception of IBD, in which the MHC association is weaker overall). However, independent MHC loci can now explain a comparable amount of the disease variance to that explained by the non-MHC associated genes known so far.

### Insights into MHC susceptibility for infectious diseases: GWAS, fine-mapping results, and epistasis

In principle, an infectious disease is caused by interactions between a pathogen, the environment, and host genetics. Here, we discuss MHC genetic associations reported in infectious diseases from GWAS (Table 2) and how these findings can explain increased susceptibility or protection by affecting human immune responses. This is why certain MHC classes are important in infectious diseases. We note that fewer MHC associations have been found for infectious diseases than for autoimmune diseases, mainly because of the smaller cohort sizes for infectious diseases. Thus, extensive fine-mapping studies (and imputation) have yet to be performed, with the exception of a few studies on infections such as human immunodeficiency virus (HIV) [62], human hepatitis B virus (HBV) [63, 64], human hepatitis C virus (HCV) [65], human papilloma virus (HPV) seropositivity [66], and tuberculosis [67].

From a genetic viewpoint, one of the best-studied infectious diseases is HIV infection. MHC class I loci have strong effects on HIV control [62, 68–71] and acquisition [72], viral load set point [69–71], and non-progression of disease [73] in Europeans [69, 70, 72, 73], and in multi-ethnic populations (Europeans, African-Americans, Hispanics, and Chinese) [62, 68, 71]. A GWAS of an African-American population indicated a similar HIV-1 mechanism in Europeans and African-Americans: about 9.6% of the observed variation in viral load set point can be explained by HLA-B*5701 in Europeans [69], while about 10% can be explained by HLA-B*5703 in African-Americans [68]. In contrast, the MHC associations and imputed amino acids identified in Europeans and African-Americans were not replicated in Chinese populations, possibly because of the varied or low minor allele frequencies of these SNPs in Chinese people [71]. A strong association to the MHC class I polypeptide-related sequence B (*MICB*) was also revealed by a recent GWAS for dengue shock syndrome (DSS) in Vietnamese children [74]. This result was replicated in Thai patients, indicating *MICB* can be a strong risk factor for DSS in Southeastern Asians [75].

HLA-DP and HLA-DQ loci, along with other MHC or non-MHC loci (TCF19, EHMT2, HLA-C, HLA-DOA, UBE2L3, CFB, CD40, and NOTCH4) are consistently associated with susceptibility to HBV infection in Asian populations [76–83]. Significant associations between the HLA-DPA1 locus and HBV clearance were also confirmed in independent East Asian populations [79, 81]. A fine-mapping study of existing GWAS data from Han Chinese patients with chronic HBV infection used SNP2HLA as the imputation tool and a pan-Asian reference panel. It revealed four independent associations at

**Table 2** Major histocompatibility complex (MHC) associations and risks for infectious diseases identified by genome-wide association studies (GWAS)

| Phenotype | Initial sample size | Replication sample size | Mapped / nearest gene | SNPs | *P*-value | OR or beta | Reference |
|---|---|---|---|---|---|---|---|
| HIV-1 viral set point | Euro-CHAVI, 486 seroconverters | 140 Caucasian patients | HCP5/B*5701 | rs2395029 | $9.36 \times 10^{-12}$ | NA | [69] |
| | | | HLA-C | rs9264942 | $3.77 \times 10^{-9}$ | NA | |
| HIV-1 control | 2362 European ancestry cases | NA | HSPA1B-C6orf48 | rs9368699 | $5.00 \times 10^{-08}$ | NA | [70] |
| | | | LOC105375015 | rs9264942 | $6.00 \times 10^{-32}$ | NA | |
| | | | LOC105375015 | rs9264942 | $6.00 \times 10^{-12}$ | NA | |
| | | | HCP5 | rs2395029 | $5.00 \times 10^{-35}$ | NA | |
| | | | HCP5 | rs2395029 | $1.00 \times 10^{-11}$ | NA | |
| | | | PSORS1C1 | rs3815087 | $8.00 \times 10^{-08}$ | NA | |
| AIDS progression | 275 seropositive non-progressors, 86 seropositive rapid progressors, 1352 seronegative controls, All Europeans | See [69] | HCP5 | rs2395029 | $3.00 \times 10^{-19}$ | | [73] |
| | | | LOC105375015 | rs10484554 | $6.00 \times 10^{-08}$ | NA | |
| HIV-1 control | 974 controllers (cases) and 2648 progressors (controls) multi ethnic (Europeans, African-Americans, and Hispanics) | Untreated HIV-infected persons from Switzerland | HLA-C | rs9264942 | $2.8 \times 10^{-35}$ | 2.9 | [62] |
| | | | HLA-B*57:01 | rs2395029 | $9.7 \times 10^{-26}$ | 5.3 | |
| | | | MICA | rs4418214 | $1.4 \times 10^{-34}$ | 4.4 | |
| | | | PSORS1C3 | rs3131018 | $4.2 \times 10^{-16}$ | 2.1 | |
| | | | HLA-B | rs2523608 | $8.9 \times 10^{-20}$ | 2.6 | |
| | | | Intergenic | rs2255221 | $3.5 \times 10^{-14}$ | 2.7 | |
| | | | HLA-B | rs2523590 | $1.7 \times 10^{-13}$ | 2.4 | |
| HIV-1 susceptibility | 6334 cases, 7247 controls (Europeans) | NA | MICA - LOC105375017 | rs4418214 | $4.00 \times 10^{-11}$ | 1.52 | [72] |
| Chronic hepatitis B infection | 786 Japanese cases, 2201 Japanese controls | Japanese and Thai cohorts of 1300 cases and 2100 controls | HLA-DPA1 | rs3077 | $2.31 \times 10^{-38}$ | 0.56 | [80] |
| | | | HLA-DPB1 | rs9277535 | $6.34 \times 10^{-39}$ | 0.57 | |
| Chronic hepatitis B infection | 458 Japanese cases, 2056 Japanese controls | Three independent Japanese cohorts (2209 cases and 4440 controls) | HLA-DPA1 | rs3077 | $1.28 \times 10^{-61}$ | 1.98 | [83] |
| | | | HLA-DPB1 | rs9277535 | $3.72 \times 10^{-17}$ | 1.95 | |
| | | | HLA-DQB1 | rs2856718 | $4.41 \times 10^{-10}$ | 1.59 | |
| | | | HLA-DQB2 | rs7453920 | $1.27 \times 10^{-10}$ | 2.2 | |
| Chronic hepatitis B infection | 400 cases, 1000 controls, Koreans | 971 cases, 1938 controls Korean | TCF19 | rs1419881 | $1.00 \times 10^{-18}$ | 1.37 | [77] |
| | | | EHMT2 | rs652888 | $7.00 \times 10^{-13}$ | 1.38 | |
| | | | HLA-DQB1 - LOC102725019 | rs2856718 | $2.00 \times 10^{-24}$ | 1.6 | |
| | | | HLA-DQB2 | rs7453920 | $7.00 \times 10^{-26}$ | 2 | |
| | | | HLA-DPA1 | rs3077 | $5.00 \times 10^{-39}$ | 1.89 | |
| | | | HLA-DPB1 | rs9277535 | $4.00 \times 10^{-40}$ | 1.89 | |
| Hepatitis B | 951 carrier cases, 937 cleared controls, Han Chinese | 4230 carrier cases, 3051 cleared controls, 2622 controls, Han Chinese | HCG27 - HLA-C | rs3130542 | $9.00 \times 10^{-14}$ | 1.33 | [76] |
| | | | HLA-DQB2 | rs7453920 | $5.00 \times 10^{-37}$ | 1.8868 | |

**Table 2** Major histocompatibility complex (MHC) associations and risks for infectious diseases identified by genome-wide association studies (GWAS) *(Continued)*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Chronic hepatitis B infection | 321 cases, 304 controls, TaiwaneseHan Chinese | 1302 cases, 761 controls, Taiwanese Han Chinese | LOC102725019 - HLA-DQA2 | rs9276370 | $2.00 \times 10^{-12}$ | 1.95 | |
| | | | HLA-DQB2 | rs7453920 | $7.00 \times 10^{-15}$ | 2.31 | [79] |
| | | | HLA-DPB1 | rs9277535 | $5.00 \times 10^{-14}$ | 1.59 | |
| | | | LOC105375021 | rs9366816 | $3.00 \times 10^{-10}$ | 1.43 | |
| | | | HLA-DQB1 - HLA-DQA2; LOC102725019 - HLA-DQA2; LOC102725019 - HLA-DQA2; LOC102725019 - HLA-DQA2; HLA-DQB1 - HLA-DQA2; LOC102725019 - HLA-DQA2; LOC102725019 - HLA-DQA2; LOC102725019 - HLA-DQA2, HLA-DQB2; HLA-DQB2; HLA-DQB2; HLA-DQB2; HLA-DQB2; HLA-DQB2; HLA-DQB2; HLA-DQB2, HLA-DPB1; HLA-DPB1; HLA-DPB1; HLA-DPB1; HLA-DPB1; HLA-DPB1; HLA-DPB1; HLA-DPB1, LOC105375021; LOC105375021 | rs9276370, rs7756516, rs7453920, rs9277535, rs9366816 | $1.00 \times 10^{-12}$ | 1.49 | |
| Chronic hepatitis B infection | 2514 cases, 1130 controls, Chinese | 6600 cases, 8127 controls, Chinese | HCG27 - HLA-C | rs2853953 | $5.00 \times 10^{-20}$ | 1.47 | [78] |
| | | | CFB | rs12614 | $1.00 \times 10^{-34}$ | 1.89 | |
| | | | NOTCH4 | rs422951 | $5.00 \times 10^{-16}$ | 1.27 | |
| | | | HLA-DQB1 - LOC102725019 | rs2856718 | $7.00 \times 10^{-28}$ | 1.28 | |
| | | | HLA-DQB2 | rs7453920 | $1.00 \times 10^{-60}$ | 2 | |
| | | | HLA-DOA | rs378352 | $1.00 \times 10^{-23}$ | 1.26 | |
| | | | HLA-DPA1 | rs3077 | $1.00 \times 10^{-53}$ | 1.45 | |
| | | | HLA-DPB1 | rs9277535 | $1.00 \times 10^{-70}$ | 1.52 | |
| Chronic hepatitis B infection/progression | 181 Japanese carriers, 184 Japanese healthy controls | 256 carriers, 236 healthy controls (Japanese) and 344 carriers, 151 healthy controls (Korean) | 3' UTR of HLA-DPB1 | rs9277542 | $8.36 \times 10^{-8}$ | 0.42 | [81] |
| Chronic hepatitis C infection | 1482 chronic cases, 919 spontaneously cleared cases, European, African, and mixed/ other ancestry | 461 chronic cases, 284 spontaneously cleared cases | HLA-DQB1 - LOC102725019 | rs4273729 | $5.00 \times 10^{-17}$ | 1.59 | [65] |
| Chronic hepatitis C infection | 481 cases, 2963 controls, Japanese | 5737 cases, 26,931 controls Japanese | HLA-DQB1 - LOC102725019 | rs9275572 | $4.00 \times 10^{-16}$ | 1.27 | [84] |
| HPV seropositivity | 1286 lung cancer cases, 679 head and neck cancer cases, 811 kidney cancer cases, 2035 controls, European ancestry | 1307 head and neck cancer cases, 1037 controls, Hispanic | HLA-DQB1 - LOC102725019 | rs9357152 | $1.00 \times 10^{-14}$ | 2.02 | [85] |

**Table 2** Major histocompatibility complex (MHC) associations and risks for infectious diseases identified by genome-wide association studies (GWAS) *(Continued)*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Dengue shock syndrome | 2008 child cases, 2018 child controls, Vietnamese | 1737 child cases, 2,934 child controls, Vietnamese | MICB | rs3132468 | $4.00 \times 10^{-11}$ | 1.34 | [74] |
| Leprosy | 706 cases, 1225 controls Han Chinese | 3254 cases, 5955 controls, Chinese | HLA-DRB1 - HLA-DQA1 | rs602875 | $5.00 \times 10^{-27}$ | 1.61 | [88] |
| Leprosy | 1548 cases, 2150 controls, 4362 controls with immune-related diseases, Chinese ancestry | 6765 cases, 9505 controls, Chinese ancestry | HLA-DRB1 - HLA-DQA1 | rs9271100 | $8.00 \times 10^{-95}$ | 1.68 | [87] |
| *M. Tuberculosis* infection | 8162 cases, 277,643 controls, Icelanders | 5530 Russian cases, 5607 Russian controls and 438 cases from Croatia, 1009 controls from Croatia | HLA-DQA1, HLA-DRB1 | rs557011 | $3.1 \times 10^{-13}$ | 1.14 | [67] |
| Pulmonary tuberculosis | | | HLA-DQA1, HLA-DRB1 | rs557011 | $5.8 \times 10^{-12}$ | 1.25 | |
| Pulmonary tuberculosis | | | HLA-DQA1, HLA-DRB1 | rs9271378 | $2.5 \times 10^{-12}$ | 0.78 | |
| *M. Tuberculosis* infection | | | HLA-DQA1 | rs9272785 | $9.3 \times 10^{-9}$ | 1.14 | |
| Leishmaniasis (visceral) | 989 cases, 1089 controls (Indians), 357 cases, 1613 unaffected relatives (Brazilians) | 951 cases, 990 controls, Indians | HLA-DRB1 - HLA-DQA1 | rs9271858 | $3.00E^{-17}$ | 1.41 | [92] |

Only genome-wide significant single nucleotide polymorphisms (SNPs) are given ($P > 5 \times 10^{-8}$), which are located in the MHC region (chr6:28,477,797-33,448,354). Assembly GRCh37/hg19
*AIDS* acquired immune deficiency syndrome, *NA* not applicable, *OR* odds ratio

HLA-DPβ1 positions 84–87, HLA-C amino acid position 15, rs400488 at HCG9, and HLA-DRB1*13; together, these four associations could explain over 72.94% of the phenotypic variance caused by genetic variations [64]. Another recent study using imputed data from Japanese individuals indicated that class II alleles were more strongly associated with chronic HBV infection than class I alleles (Additional file 1) [63]. Similarly, the HLA-DQ locus influences the spontaneous clearance of HCV infection in cohorts of European and African ancestry, while DQB1*03:01, which was identified by HLA genotyping together with the non-MHC IL28B, can explain 15% of spontaneous HCV infection clearance cases [65]. HLA-DQB1*03 also confers susceptibility to chronic HCV in Japanese people [84]. A GWAS in a European population revealed that HPV8 seropositivity is influenced by the MHC class II region [85]. However, HPV type 8 showed a higher seropositivity prevalence than other HPV types at the population level [66]; this led to a limited power to detect associations with other HPV types. Fine-mapping using the same European population as in the GWAS [66] revealed significant associations with HPV8 and HPV77 seropositivity, but only with MHC class II alleles, not with class I alleles. This indicates a pivotal role for class II molecules in antibody immune responses in HPV infection. Notably in this study, imputation was performed using HLA*IMP:02 and reference panels from the HapMap Project [86] and the 1958 British Birth Cohort, as well as using SNP2HLA with another reference panel from the T1DGC. Both imputation tools provided comparable results, thus highlighting the important role of MHC class II alleles in antibody response to HPV infection [66].

A GWAS on leprosy in Chinese populations pointed to significant associations with HLA-DR-DQ loci [87, 88]; these results were replicated in an Indian population [89]. Fine-mapping the MHC showed that variants in HLA class II were extensively associated with susceptibility to leprosy in Chinese people, with HLA-DRB1*15 being the most significant variant [87]. HLA class II variants also influence the mycobacterial infection tuberculosis in European and African populations [67, 90]. Fine-mapping identified the DQA1*03 haplotype, which contains four missense variants and contributes to disease susceptibility [67]. A meta-analysis showed that five variants (HLA-DRB1*04, *09, *10, *15, and *16) increase the risk of tuberculosis, especially in East Asian populations, whereas HLA-DRB1*11 is protective [91].

Using a population from Brazil, the first GWAS on visceral leishmaniasis revealed that the class II HLA-DRB1-HLA-DQA1 locus had the strongest association signal; this was replicated in an independent Indian population [92]. This common association suggests that Brazilians and Indians share determining genetic factors that are independent of the different parasite species in these geographically distinct regions.

Finally, epistatic interactions between MHC class I alleles and certain KIR alleles (between KIR3DS1 combined with HLA-B alleles) are associated with slower progression to acquired immunodeficiency syndrome (AIDS) [93] and better resolution of HCV infection (between KIR2DL3 and its human leukocyte antigen C group 1, HLA-C1) [94].

#### Insights into the biology of infectious diseases

Associations with the MHC class I locus suggest a critical role for $CD8^+$ T-cell responses in major viral infections such as HIV, dengue, and HCV. This critical role of $CD8^+$ T-cell responses in HIV infection is reflected by the slow disease progression seen in infected individuals because of their increasing $CD8^+$ T-cell responses that are specific to conserved HIV proteins such as Gap p24 [95]. Interestingly, five out of six amino acid residues (Additional file 1) identified as associated with HIV control [62] lie in the MHC class I peptide-binding groove, implying that MHC variation affects peptide presentation to $CD8^+$ T cells. In particular, the amino acid at position 97, which lies in the floor of the groove in HLA-B, was most significantly associated with HIV control ($P = 4 \times 10^{-45}$) [62]. This amino acid is also implicated in MHC protein folding and cell surface expression [96]. An association found in severe dengue disease also underscores the role of $CD8^+$ T cells in disease pathogenesis: class I alleles that were associated with an increased risk of severe dengue disease were also associated with weaker $CD8^+$ T-cell responses in a Sri Lankan population from an area of hyper-endemic dengue disease [97]. In HCV, similar to the protective alleles against HIV infection [95], HLA-B*27 presents the most conserved epitopes of HCV to elicit strong cytotoxic T-cell responses, thereby reducing the ability of HCV to escape from host immune responses [98].

Associations between genetic variants in the MHC class II region and disease susceptibility imply that impaired antigen presentation or unstable MHC class II molecules contribute to insufficient $CD4^+$ T-cell responses and, subsequently, to increased susceptibility to infections. For instance, the amino acid changes at positions of HLA-DPβ1 and HLA-DRβ1 in the antigen-binding groove that influence HBV infection may result in defective antigen presentation to $CD4^+$ T cells or to impaired stability of MHC class II molecules, thereby increasing susceptibility to HBV infection [64]. $CD4^+$ T-cell responses are also critical in mycobacterial infections, such as has been described for leprosy and tuberculosis [99, 100]. Notably, monocyte-derived macrophages treated with live *Mycobacterium leprae* showed three main responses that explain infection

persistence: downregulation of certain pro-inflammatory cytokines and MHC class II molecules (HLA-DR and HLA-DQ), preferentially primed regulatory T-cell responses, and reduced Th1-type and cytotoxic T-cell function [99]. Macrophages isolated from the lesions of patients with the most severe disease form, lepromatous leprosy, also showed lower expression of MHC class II molecules, providing further evidence that defective antigen presentation by these molecules leads to more persistent and more severe *M. leprae* infection [99].

Recently, it has been shown that $CD4^+$ T-cells are essential for the optimal production of IFNγ by $CD8^+$ T-cells in the lungs of mice infected with *M. tuberculosis*, indicating that communication between these two distinct effector cell populations is critical for a protective immune response against this infection [101]. Impaired antigen processing and presentation from *Leishmania*-infected macrophages (which are the primary resident cells for this parasite) to $CD4^+$ T cells could explain increased susceptibility to leishmaniasis [102]. The association between HPV seropositivity and the MHC class II region also suggests that class II molecules bind and present exogenous antigens more effectively to a subset of $CD4^+$ T cells known as Th2. These Th2 cells help primed B lymphocytes to differentiate into plasma cells and to secrete antibodies against the HPV virus.

In support of the hypothesis that genetic effects on both $CD8^+$ (class I) and $CD4^+$ (class II) cells modify the predisposition to infections, it should be noted that some infectious diseases, such as HIV, HBV, HCV, and leprosy, show associations to more than one of the classic MHC classes and, in some cases, the associations differ between populations (Table 2). Moreover, consideration must be given to the differences between viral and bacterial genotypes in the same infection, which play a role in determining potentially protective effects. Overall, associations with multiple MHC loci reflect the complex and interactive nature of host immune responses when the host encounters a pathogen.

## Relationship between the MHC variants involved in autoimmune and infectious diseases

Both autoimmune and infectious diseases seem to involve certain MHC classes (Fig. 2a), and only a few MHC alleles are shared between these two distinct disease groups (Fig. 2b). The identification of shared MHC variation has provided insight into the relationships between the MHC variants involved in autoimmune and infectious diseases and which have been uniquely shaped throughout human evolution [18].

Two hypotheses have been proposed to explain the relationships between the MHC variants involved in both groups of diseases. The first, known as the "pathogen-driven selection" hypothesis, states that pressure exerted on the human genome by pathogens has led to the advantageous selection of host defense genes and, subsequently, to much higher polymorphism in the MHC. This polymorphism has contributed to the development of complex immune defense mechanisms that protect humans against a broad range of pathogens. Thus, heterozygosity at MHC loci is evolutionarily favored and has become an efficient mechanism contributing to the highly polymorphic MHC (the "MHC heterozygosity advantage") [103]. Two examples of MHC heterozygote advantage are HIV-1-infected heterozygotes at class I loci, which are slower to progress to AIDS [104, 105], and HBV-infected heterozygotes at class II loci, which seem more likely to clear the infection [106]. In addition, human populations exposed to a more diverse range of pathogens display higher class I genetic diversity than those exposed to a smaller range [107]. However, the true effect of infectious diseases on selection might be underestimated because of the heterogeneity of many pathogens and the changing prevalence of infectious diseases over evolutionary time.

Positive selection of the advantageous effect of MHC polymorphism in infections may also be accompanied by a higher risk of developing autoimmune diseases. For example, the non-MHC locus *SH2B3* rs3184504*A is a risk allele for CeD but has been under positive selection because it offers the human host protection against bacterial infections [108]. To investigate whether other genetic variants in the MHC show this opposite direction effect between autoimmune and infectious diseases (Fig. 2b), we compared SNPs and alleles in the MHC identified by GWAS and fine-mapping studies on autoimmune diseases (Table 1; Additional file 2) with those identified in infectious diseases (Table 2; Additional file 1). On the one hand, HLA-B*27:05, which has one of the strongest associations to AS in the MHC ($P < 1 \times 10^{-2000}$) [37] and is present in all ethnic groups, increases AS risk. On the other hand, it also has a protective effect against HIV infection, showing a nominal significant value of $5.2 \times 10^{-5}$ [70]. The second example of opposite allelic effect is the association between the rs2395029*G allele and susceptibility to psoriasis (OR = 4.1; $P = 2.13 \times 10^{-26}$) [109] and AIDS non-progression ($P = 9.36 \times 10^{-12}$) [69]. Located in the HLA complex P5 (HCP5), rs2395029 is a proxy for the HLA-B*57:01 allele [69], the strongest protective allele against AIDS progression [110]. Non-progressors carrying the rs2395029-G allele had a lower viral load than other non-progressors [73].

Another study showed that psoriasis patients carry the same genetic variants as HIV controllers/non-progressors and that they are particularly enriched for the protective allele HLA-B*57:01 ($P = 5.50 \times 10^{-42}$) [111]. Moreover, the intergenic variant rs10484554*A, which is

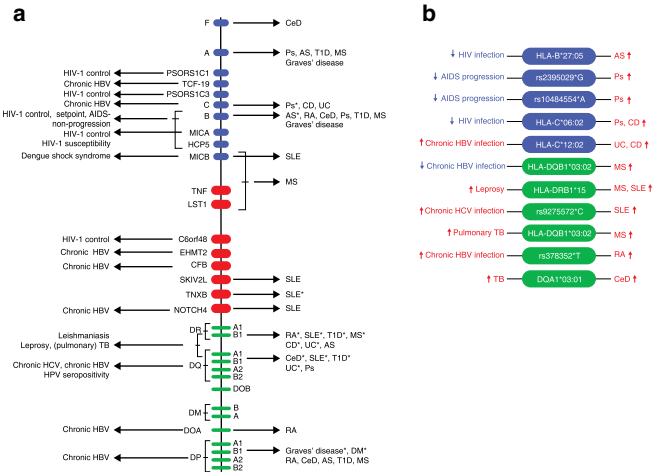



**Fig. 2** Major histocompatibility complex allele associations with autoimmune and infectious diseases. **a** Abbreviations marked with an asterisk indicate the autoimmune disease showing the strongest association with the specific locus. **b** Single nucleotide polymorphisms (SNPs) and alleles in the major histocompatibility complex (MHC) shared between autoimmune and infectious diseases. The *blue area* shows MHC alleles located in the class I region and the *green area* shows those in the class II region. The *blue arrows* indicate either a protective effect of the genetic variants against the infectious disease or a slower progression to the infectious disease. The *red arrows* indicate increased susceptibility to the corresponding autoimmune or infectious disease. *AIDS* acquired immunodeficiency syndrome, *AS* ankylosing spondylitis, *CD* Crohn's disease, *CeD* celiac disease, *DM* dermatomyositis, *HBV* hepatitis B virus, *HCV* hepatitis C virus, *HIV* human immunodeficiency virus, *MS* multiple sclerosis, *Ps* psoriasis, *RA* rheumatoid arthritis, *SLE* systemic lupus erythematosus, *T1D* type 1 diabetes, *TB* tuberculosis, *UC* ulcerative colitis, *HPV* human papilloma virus

in LD with HLA-C ($r^2 \geq 0.8$), was significantly associated with AIDS non-progression ($P = 6.27 \times 10^{-8}$) [73] and with susceptibility to psoriasis (OR = 4.66, $P = 4 \times 10^{-214}$) [58]. HLA-C*06:02 (equivalent to HLA-Cw6) was most strongly associated with susceptibility to psoriasis (OR = 3.26; $P = 2.1 \times 10^{-201}$) [36] and is also protective against HIV infection (OR = 2.97; $P = 2.1 \times 10^{-19}$) [62]. The same allele has been associated with susceptibility to CD (OR = 1.17; $P = 2 \times 10^{-13}$) [46]. Interestingly, the role of MHC in HIV control also relates to the influence of MHC expression levels. For instance, rs9264942 shows one of the most significant genome-wide effects observed on HIV control [62, 69, 70]: it is located 35 kb upstream of the HLA-C locus (Table 2) and has been associated with high HLA-C expression, conferring protection against HIV infection [112]. Explaining this protective effect, HLA-C allelic expression was correlated with increasing likelihood of $CD8^+$ T-cell cytotoxicity [112]. However, the −35 SNP is not a causal variant, but is in LD with a SNP at the 3′ end of HLA-C; this affects HLA-C expression by influencing binding of the microRNA Hsa-miR-148a [113]. Notably, high HLA-C expression has a deleterious effect by conferring risk for Crohn's disease [113]. The potential mechanism by which HLA expression levels confer resistance to pathogens, and also lead to greater autoimmunity, could be through promiscuous peptide binding [114]. Lastly, HLA-DQB1*03:02 showed a dominant risk effect for MS (OR = 1.30; $P = 1.8 \times 10^{-22}$) [45], whereas it is a resistant allele against chronic HBV infection (OR = 0.59; $P = 1.42 \times 10^{-5}$) [63].

The second hypothesis states that pathogens can trigger autoimmunity, as suggested by epidemiological studies [115, 116]. For example, it has recently been shown that apoptosis of infected colonic epithelial cells in mice induces the proliferation of self-reactive $CD4^+$ T cells that are specific to cellular and to pathogenic antigens [117]. Self-reactive $CD4^+$ T cells differentiate into Th17 cells, which promote production of auto-antibodies and auto-inflammation, implying that infections can trigger autoimmunity [117]. Other mechanisms have been proposed, such as molecular mimicry, bystander activation, exposure of cryptic antigens, and superantigens [118]. Common genetic signatures between autoimmune and infectious diseases indirectly imply that pathogens can indeed trigger autoimmunity. In line with this second hypothesis, we have identified common risk factors between autoimmune and infectious diseases, such as the

alleles: HLA-DRB1*15 for MS, SLE (Table 1), and leprosy (OR = 2.11; $P = 3.5 \times 10^{-28}$) [87]; rs9275572*C, located in HLA-DQ, for chronic HCV infection (OR = 0.71; $P = 2.62 \times 10^{-6}$) [84], and SLE ($P = 1.94 \times 10^{-6}$) [119]; HLA-DQB1*03:02 for MS (OR = 1.30; $P = 1.8 \times 10^{-22}$) [45] and pulmonary tuberculosis (OR = 0.59; $P = 2.48 \times 10^{-5}$) [67]; HLA-C*12:02 for UC (OR = 2.25; $P = 4 \times 10^{-37}$) [46], CD (OR = 1.44; $P = 3x\ 10^{-8}$) [46], and chronic HBV infection (OR = 1.70; $P = 7.79 \times 10^{-12}$) [63]; and rs378352*T, located in HLA-DOA, for chronic HBV infection (OR = 1.32; P = $1.16 \times 10^{-7}$) [78] and RA (OR = 1.24; $P = 4.6 \times 10^{-6}$) [25] (Fig. 2a).

Associations within the MHC region for several autoimmune diseases, such as RA, CeD, AS, T1D, Graves' disease, and DM, and HBV infection are driven by variants and alleles around HLA-DPB1 (Table 1), implying that viruses like HBV could trigger autoimmunity. Although there is no convincing evidence, HBV and HCV are associated with extra-hepatic autoimmune perturbations [120, 121]. Lastly, the DQA1*03:01 allele, which contributes to tuberculosis susceptibility (OR = 1.31; $P = 3.1 \times 10^{-8}$) [67], is also a well-known risk factor for CeD as part of the DQ8 (DQA1*03-DQB1*03:02) and DQ2.3 (trans-DQA1*03:01 and DQB1*02:01) haplotypes [122]. DQA1*03 also increases susceptibility to T1D, RA, and juvenile myositis [123–125]. Overall, the direction of association is the same for most shared MHC class II loci, suggesting that bacteria and viruses can trigger immune responses. No viruses have been proven to cause an autoimmune disease thus far, but multiple virus infections could prime the immune system and eventually trigger an autoimmune response; this is a hypothesis that has been supported by animal studies on MS [126].

## Conclusions and future perspectives

We have discussed recent advances in understanding the genetic variation in the MHC in relation to autoimmune and infectious diseases. However, confidence in the associations between MHC and infectious diseases is limited, mainly because of the relatively small patient cohort sizes available. Further limitations to identifying and replicating associations with infectious diseases include: strain differences, heterogeneity in clinical phenotypes, use of inappropriate controls (such as individuals with asymptomatic infections), and population-specific differences in allele frequency and/or haplotype structure. Finally, with the exception of a few described above, no imputation has been performed in most infectious disease studies. In certain populations, such as Africans, lower LD makes it challenging to perform MHC imputation.

Although application of a traditional GWAS is challenging for infectious diseases, other approaches may increase the power of genetic studies. For instance, a combination of transcriptional analysis and systems biology allowed the identification of a novel role for type I IFN signaling pathway in the human host immune response against *Candida albicans* [127]. The use of control subjects for whom it is known whether they clear the infection, and who come from the same hospital as patients, could be appropriate for infectious diseases so that co-morbidities and clinical risk factors are as similar as possible between groups. Overall, initiating collaborative efforts to increase patient cohort numbers, designing better studies by using more appropriate controls and more homogenously clinically defined patient phenotypes, and applying imputation using population-specific reference genomes would open new avenues to study the genetics of infectious diseases.

In contrast to infectious diseases, the added value of fine-mapping the MHC to pinpoint genetic risk factors for autoimmune disease has been well demonstrated by numerous studies. The associations that have been found in both European and Asian populations to the same amino acids by fine-mapping the MHC suggest that the same molecular mechanism is involved, despite the differences in MHC allele frequencies seen between these ethnic groups.

MHC-based imputation approaches using genotype data, along with the use of population-specific reference panels for imputing MHC alleles and amino acids, has allowed identification of the MHC variation associated with complex diseases. Although identification is challenging, genetic variation in the MHC is of critical importance for two reasons. First, it sheds light on the development of autoimmunity, given the two hypotheses discussed above (pathogen-driven evolutionary selection of protective genes or pathogens as triggers of autoimmunity), and second, it yields greater understanding of the complexity of the human immune system. This knowledge will ultimately permit the design of better prophylactic and therapeutic strategies to achieve more balanced patient–immune responses during treatment.

## Additional files

**Additional file 1:** Imputed MHC classic alleles and amino acids for infectious diseases. (XLSX 40 kb)

**Additional file 2:** GWAS SNPs for autoimmune diseases. (XLSX 61 kb)

### Abbreviations

APC: Antigen-presenting cell; AS: Ankylosing spondylitis; CD: Crohn's disease; CeD: Celiac disease; DM: Dermatomyositis; GWAS: Genome-wide association study; HBV: Hepatitis B virus; HCV: Hepatitis C virus; HIV: Human immunodeficiency virus; HLA: Human leukocyte antigen; HPV: Human papilloma virus; IBD: Inflammatory bowel disease; KIR: Killer immunoglobulin receptor; LD: Linkage disequilibrium; MHC: Major histocompatibility complex; MS: Multiple sclerosis; NGS: Next-generation sequencing; PRG: Population reference graph; RA: Rheumatoid arthritis; SLE: Systemic lupus erythematosus; SNP: Single nucleotide polymorphism; TCR: T-cell receptor; UC: Ulcerative colitis

**Acknowledgements**
We thank Kate McIntyre and Jackie Senior for editing the manuscript and Aggelo Matzaraki for help with graphic design.

**Funding**
VM is supported by a PhD scholarship from the Graduate School of Medical Sciences, University of Groningen, the Netherlands. VK is supported by the European Union's Framework Programme 7 (EU FP7) TANDEM project (HEALTH-F3-2012-305279). CW is supported by an FP7/2007-2013/ERC Advanced Grant (agreement 2012-322698), the Stiftelsen KG Jebsen Coeliac Disease Research Centre (Oslo, Norway), and a Spinoza Prize from the Netherlands Organization for Scientific Research (NWO SPI 92-266). AZ holds a Rosalind Franklin Fellowship (University of Groningen) and is also supported by CardioVasculair Onderzoek Nederland (CVON 2012-03).

**Authors' contributions**
VM and AZ drafted the manuscript. VK and CW critically revised the manuscript. All authors read and approved the final manuscript.

**Competing interests**
The authors declare that they have no competing interests.

## Publisher's Note



Published online: 27 April 2017

**References**

1. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet. 2009;54:15–39.
2. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;10:8192.
3. Song S, Miranda CJ, Braun L, Meyer K, Frakes AE, Ferraiuolo L, et al. Major histocompatibility compex class I molecules protect motor neurons from astrocyte-induced toxicity in amyotrophic lateral sclerosis (ALS). Nat Med. 2016;22:397–403.
4. Shatz CJ. MHC class I: an unexpected role in neuronal plasticity. Neuron. 2009;64(1):40–5.
5. Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet. 2010;42:781–5.
6. Hill-Burns EM, Factor SA, Zabetian CP, Thomson G, Payami H. Evidence for more than one Parkinson's disease-associated variant within the HLA region. PLoS One. 2011;6, e27109.
7. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011;13:101.
8. Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, Almeida J, et al. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics. 2008;60:1–18.
9. Hughes T, Adler A, Kelly JA, Kaufman KM, Williams AH, Langefeld CD, et al. Evidence for gene-gene epistatic interactions among susceptibility loci for systemic lupus erythematosus. Arthritis Rheum. 2012;64:485–92.
10. Kirino Y, Bertsias G, Ishigatsubo Y, Mizuki N, Tugal-Tutkun I, Seyahi E, et al. Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. Nat Genet. 2013;45:202–7.
11. Carapito R, Radosavljevic M, Bahram S. Next-generation sequencing of the HLA locus: methods and impacts on HLA typing, population genetics and disease association studies. Hum Immunol. 2016;77(11):1016–23.
12. Nagasaki M, Yasuda J, Katsuoka F, Nariai N, Kojima K, Kawai Y, et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat Commun. 2015;6:8018.
13. Castelli EC, Mendes-Junior CT, Sabbagh A, Porto IO, Garcia A, Ramalho J, et al. HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. Hum Immunol. 2015;76:945–53.
14. Rhoads A, Au KF. PacBio sequencing and its applications. Genomics Proteomics Bioinforma. 2015;13:278–89.
15. Gowda M, Ambardar S, Dighe N, Manjunath A, Shankaralingu C, Hallappa P, et al. Comparative analyses of low, medium and high-resolution HLA typing technologies for human populations. J Clin Cell Immunol. 2016;7:399.
16. Zhou F, Cao H, Zuo X, Zhang T, Zhang X, Liu X, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. Nat Genet. 2016;48:1–11.
17. Cho JH, Feldman M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. Nat Med. 2015;21:730–8.
18. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet. 2010;11:17–30.
19. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee H, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012;44:291–6.
20. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One. 2013;8(6), e64683.
21. Dilthey AT, Moutsianas L, Leslie S, McVean G. HLA*IMP–an integrated framework for imputing classical HLA alleles from SNP genotypes. Bioinformatics. 2011;27:968–72.
22. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, Nelson MR, et al. Multi-population classical HLA type imputation. PLoS Comput Biol. 2013;9(2), e1002877.
23. Nariai N, Kojima K, Saito S, Mimori T, Sato Y, Kawai Y, et al. HLA-VBSeq: accurate HLA typing at full resolution from whole-genome sequencing data. BMC Genomics. 2015;16 Suppl 2:S7.
24. Okada Y, Momozawa Y, Ashikawa K, Kanai M, Matsuda K, Kamatani Y, et al. Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. Nat Genet. 2015;47:798–802.
25. Okada Y, Suzuki A, Ikari K, Terao C, Kochi Y, Ohmura K, et al. Contribution of a non-classical HLA gene, HLA-DOA, to the risk of rheumatoid arthritis. Am J Hum Genet. 2016;99:366–74.
26. Kim K, Bang SY, Lee HS, Bae SC. Construction and application of a Korean reference panel for imputing classical alleles and amino acids of human leukocyte antigen genes. PLoS One. 2014;9(11), e112546.
27. Dilthey A, Cox C, Iqbal Z, Nelson MR, McVean G. Improved genome inference in the MHC using a population reference graph. Nat Genet. 2015;47:682–8.
28. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526:68–74.
29. Lenz TL, Deutsch AJ, Han B, Hu X, Okada Y, Eyre S, et al. Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. Nat Genet. 2015;47:4–7.
30. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11:499–511.
31. Mulder DJ. HLA antigens and coeliac disease. Lancet. 1974;2:727.
32. Grumet FC, Coukell A, Bodmer JG, Bodmer WF, McDevitt HO. Histocompatibility (HL-A) antigens associated with systemic lupus erythematosus. A possible genetic predisposition to disease. N Engl J Med. 1971;285:193–6.
33. Kim K, Bang S-Y, Yoo DH, Cho S-K, Choi C-B, Sung Y-K, et al. Imputing variants in HLA-DR beta genes reveals that HLA-DRB1 is solely associated with rheumatoid arthritis and systemic lupus erythematosus. PLoS ONE. 2016;11(2):e0150283.
34. Okada Y, Kim K, Han B, Pillai NE, Ong RT-H, Saw W-Y, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Hum Mol Genet. 2014;23:6916–26.
35. Gutierrez-Achury J, Zhernakova A, Pulit SL, Trynka G. Fine-mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. Nat Genet. 2015;47:577–8.
36. Okada Y, Han B, Tsoi LC, Stuart PE, Ellinghaus E, Tejasvi T, et al. Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. Am J Hum Genet. 2014;95:162–72.
37. Cortes A, Pulit SL, Leo PJ, Pointon JJ, Robinson PC, Weisman MH, et al. Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1. Nat Commun. 2015;6:7146.

38. Kim K, Bang SY, Lee HS, Okada Y, Han B, Saw WY, et al. The HLA-DRβ1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. Nat Commun. 2014;5:5902.
39. Fernando MM, Stevens CR, Sabeti PC, Walsh EC, McWhinnie AJ, Shah A, et al. Identification of two independent risk factors for lupus within the MHC in United Kingdom families. PLoS Genet. 2007;3, e192.
40. Morris DL, Taylor KE, Fernando MMA, Nititham J, Alarcón-Riquelme ME, Barcellos LF, et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. Am J Hum Genet. 2012;91:778–93.
41. Sun C, Molineros JE, Looger LL, Zhou X, Okada Y, Ma J, et al. High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. Nat Genet. 2016;48:323–30.
42. Howson JMM, Walker NM, Clayton D, Todd JA. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. Diabetes Obes Metab. 2009;11:31–45.
43. Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen W-M, et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. Nat Genet. 2015; 47:898–905.
44. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 2015;47:1107–13.
45. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, et al. Fine-mapping the genetic gssociation of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet. 2013; 9(11), e1003926.
46. Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat Genet. 2015;47:172–9.
47. Zhang CE, Li Y, Wang ZX, Gao JP, Zhang XG, Zuo XB, et al. Variation at HLA-DPB1 is associated with dermatomyositis in Chinese population. J Dermatol. 2016;43:1307–13.
48. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum. 1987;30:1205–13.
49. Koeleman BP, Lie BA, Undlien DE, Dudbridge F, Thorsby E, de Vries RR, et al. Genotype effects and epistasis in type 1 diabetes and HLA-DQ trans dimer associations with disease. Genes Immun. 2004;5:381–8.
50. van Gaalen FA, Verduijn W, Roelen DL, Böhringer S, Huizinga TWJ, van der Heijde DM, et al. Epistasis between two HLA antigens defines a subset of individuals at a very high risk for ankylosing spondylitis. Ann Rheum Dis. 2013;72:974–8.
51. Mizuki M, Eklund A, Grunewald J. Altered expression of natural killer cell inhibitory receptors (KIRs) on T cells in bronchoalveolar lavage fluid and peripheral blood of sarcoidosis patients. Sarcoidosis Vasc Diffus Lung Dis. 2000;17:54–9.
52. Momot T, Koch S, Hunzelmann N, Krieg T, Ulbricht K, Schmidt RE, et al. Association of killer cell immunoglobulin-like receptors with scleroderma. Arthritis Rheum. 2004;50:1561–5.
53. Nelson GW, Martin MP, Gladman D, Wade J, Trowsdale J, Carrington M. Cutting edge: heterozygote advantage in autoimmune disease: hierarchy of protection/susceptibility conferred by HLA and killer Ig-like receptor combinations in psoriatic arthritis. J Immunol. 2004;173:4273–6.
54. van der Slik AR, Koeleman BP, Verduijn W, Bruining GJ, Roep BO, Giphart MJ. KIR in type 1 diabetes: disparate distribution of activating and inhibitory natural killer cell receptors in patients versus HLA-matched control subjects. Diabetes. 2003;52:2639–42.
55. Moretta L, Moretta A. Killer immunoglobulin-like receptors. Curr Opin Immunol. 2004;16(5):626–33.
56. van Heemst J, Jansen DT, Polydorides S, Moustakas AK, Bax M, Feitsma AL, et al. Crossreactivity to vinculin and microbes provides a molecular basis for HLA-based protection against rheumatoid arthritis. Nat Commun. 2015;6: 6681.
57. Sheridan BS, Lefrançois L. Intraepithelial lymphocytes: to serve and protect. Curr Gastroenterol Rep. 2010;12(6):513–21.
58. Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2, Strange A, Capon F, Spencer CC, Knight J, Weale ME, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet. 2010;42:985–90.
59. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era–concepts and misconceptions. Nat Rev Genet. 2008;9:255–66.
60. Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen WM, et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. Nat Genet. 2016;47:898–905.
61. Satsangi J, Welsh KI, Bunce M, Julier C, Farrant JM, Bell JI, et al. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. Lancet. 1996;347:1212–7.
62. International HIV Controllers Study, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science. 2011;330:1551–7.
63. Nishida N, Ohashi J, Khor S, Sugiyama M, Tsuchiura T. Understanding of HLA-conferred susceptibility to chronic hepatitis B infection requires HLA genotyping- based association analysis. Sci Rep. 2016;6: 24767.
64. Zhu M, Dai J, Wang C, Wang Y, Qin N, Ma H, et al. Fine mapping the MHC region identified four independent variants modifying susceptibility to chronic hepatitis B in han chinese. Hum Mol Genet. 2015;25:1225–32.
65. Duggal P, Thio CL, Wojcik GL, Goedert JJ, Mangia A, Latanich R, et al. Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. Ann Intern Med. 2013;158: 235–45.
66. Chen D, Gaborieau V, Zhao Y, Chabrier A, Wang H, Waterboer T, et al. A systematic investigation of the contribution of genetic variation within the MHC region to HPV seropositivity. Hum Mol Genet. 2015;24: 2681–8.
67. Sveinbjornsson G, Gudbjartsson DF, Halldorsson BV, Kristinsson KG, Gottfredsson M, Barrett JC, et al. HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. Nat Genet. 2016;48:318–22.
68. Pelak K, Goldstein DB, Walley NM, Fellay J, Ge D, Shianna KV, et al. Host determinants of HIV-1 control in African Americans. J Infect Dis. 2010;201: 1141–9.
69. Fellay J, Shianna KV, Ge D, Colombo S, Weale M, Zhang K, et al. A whole-genome association study of major determinants for host control of HIV-1. Science. 2007;317:944–7.
70. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, Cirulli ET, et al. Common genetic variation and the control of HIV-1 in humans. PLoS Genet. 2009;5, e1000791.
71. Wei Z, Liu Y, Xu H, Tang K, Wu H, Lu L, et al. Genome-wide association studies of HIV-1 host control in ethnically diverse Chinese populations. Sci Rep. 2015;5:10879.
72. McLaren PJ, Coulonges C, Ripke S, van den Berg L, Buchbinder S, Carrington M, et al. Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. PLoS Pathog. 2013;9(7), e1003515.
73. Limou S, Le Clerc S, Coulonges C, Carpentier W, Dina C, Delaneau O, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). J Infect Dis. 2009;199:419–26.
74. Khor CC, Chau TNB, Pang J, Davila S, Long HT, Ong RTH, et al. Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. Nat Genet. 2011;43:1139–41.
75. Dang T, Naka I, Sa-Ngasang A, Anantapreecha S, Chanama S, Wichukchinda N, et al. A replication study confirms the association of GWAS-identified SNPs at MICB and PLCE1 in Thai patients with dengue shock syndrome. BMC Med Genet. 2014;15:58.
76. Hu Z, Liu Y, Zhai X, Dai J, Jin G, Wang L, et al. New loci associated with chronic hepatitis B virus infection in Han Chinese. Nat Genet. 2013;45: 1499–503.
77. Kim YJ, Kim HY, Lee J-H, Yu SJ, Yoon J-H, Lee H-S, et al. A genome-wide association study identified new variants associated with the risk of chronic hepatitis B. Hum Mol Genet. 2013;22:4233–8.
78. Jiang D-K, Ma X-P, Yu H, Cao G, Ding D-L, Chen H, et al. Genetic variants in five novel loci including *CFB* and *CD40* predispose to chronic hepatitis B. Hepatology. 2015;62:118–28.
79. Chang S-W, Fann CS-J, Su W-H, Wang YC, Weng CC, Yu C-J, et al. A genome-wide association study on chronic HBV infection and its clinical progression in male Han-Taiwanese. PLoS One. 2014;9, e99724.

80. Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, Hosono N, et al. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nat Genet. 2009;41:591–5.
81. Nishida N, Sawai H, Matsuura K, Sugiyama M, Ahn SH, Park JY, et al. Genome-wide association study confirming association of HLA-DP with protection against chronic hepatitis B and viral clearance in Japanese and Korean. PLoS One. 2012;7(6), e39175.
82. An P, Winkler C, Guan L, O'Brien SJ, Zeng Z, Yu Y, et al. A common HLA-DPA1 variant is a major determinant of hepatitis B virus clearance in Han Chinese. J Infect Dis. 2011;203:943–7.
83. Mbarek H, Ochi H, Urabe Y, Kumar V, Kubo M, Hosono N, et al. A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. Hum Mol Genet. 2011;20:3884–92.
84. Miki D, Ochi H, Takahashi A, Hayes CN, Urabe Y, Abe H, et al. HLA-DQB1*03 confers susceptibility to chronic hepatitis C in Japanese: a genome-wide association study. PLoS One. 2013;8(12), e84226.
85. Chen D, McKay JD, Clifford G, Gaborieau V, Chabrier A, Waterboer T, et al. Genome-wide association study of HPV seropositivity. Hum Mol Genet. 2011;20:4714–23.
86. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449:851–61.
87. Liu H, Irwanto A, Fu X, Yu G, Yu Y, Sun Y, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. Nat Genet. 2015;47:267–71.
88. Zhang F-R, Huang W, Chen S-M, Sun L-D, Liu H, Li Y, et al. Genomewide association study of leprosy. N Engl J Med. 2009;361:2609–18.
89. Wong SH, Gochhait S, Malhotra D, Pettersson FH, Teo YY, Khor CC, et al. Leprosy and the adaptation of human toll-like receptor 1. PLoS Pathog. 2010;6:1–9.
90. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet. 2010;42:739–41.
91. Tong X, Chen L, Liu S, Yan Z, Peng S, Zhang Y, et al. Polymorphisms in HLA-DRB1 gene and the risk of tuberculosis: a meta-analysis of 31 studies. Lung. 2015;193:309–18.
92. Fakiola M, Strange A, Cordell HJ, Miller EN, Pirinen M, Su Z, et al. Common variants in the HLA-DRB1-HLA-DQA1 HLA class II region are associated with susceptibility to visceral leishmaniasis. Nat Genet. 2013;45:208–13.
93. Martin MP, Gao X, Lee J-H, Nelson GW, Detels R, Goedert JJ, et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet. 2002;31:429–34.
94. Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X, Astemborski J, et al. HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. Science. 2004;305:872–4.
95. Borghans JAM, Mølgaard A, de Boer RJ, Keşmir C. HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. PLoS One. 2007;2, e920.
96. Blanco-Gelaz MA, Suárez-Alvarez B, González S, López-Vázquez A, Martínez-Borra J, López-Larrea C. The amino acid at position 97 is involved in folding and surface expression of HLA-B27. Int Immunol. 2006;18:211–20.
97. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. Proc Natl Acad Sci U S A. 2013;110:E2046–53.
98. Rao X, Hoof I, van Baarle D, Kesmir C, Textor J. HLA preferences for conserved epitopes: a potential mechanism for hepatitis C clearance. Front Immunol. 2015;6:1–9.
99. Yang D, Shui T, Miranda JW, Gilson DJ, Song Z, Chen J, et al. *Mycobacterium lepraе*-infected macrophages preferentially primed regulatory T cell responses and was associated with lepromatous leprosy. PLoS Negl Trop Dis. 2016;10:1–13.
100. Winslow GM, Cooper A, Reiley W, Chatterjee M, Woodland DL. Early T-cell responses in tuberculosis immunity. Immunol Rev. 2008;225:284–99.
101. Bold TD, Ernst JD. CD4+ T cell-dependent IFN-γ production by CD8+ effector T cells in *Mycobacterium tuberculosis* infection. J Immunol. 2012; 189:2530–6.
102. Liu D, Uzonna JE. The early interaction of Leishmania with macrophages and dendritic cells and its influence on the host immune response. Front Cell Infect Microbiol. 2012;2:83.
103. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool. 2005;2:16.
104. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, et al. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science. 1999;283:1748–52.
105. Leslie A, Matthews PC, Listgarten J, Carlson JM, Kadie C, Ndung'u T, et al. Additive contribution of HLA Class I alleles in the immune control of HIV-1 infection. J Virol. 2010;84:9879–88.
106. Thursz MR, Thomas HC, Greenwood BM, Hill AVS. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. Nat Genet. 1997;17:11–2.
107. Prugnolle F, Manica A, Charpentier M, Guégan JF, Guernier V, Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol. 2005;15:1022–7.
108. Zhernakova A, Elbers CC, Ferwerda B, Romanos J, Trynka G, Dubois PC, et al. Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. Am J Hum Genet. 2010;86:970–7.
109. Liu Y, Helms C, Liao W, Zaba LC, Duan S, Gardner J, et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS Genet. 2008;4(3), e1000041.
110. Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, et al. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proc Natl Acad Sci U S A. 2000;97:2709–14.
111. Chen H, Hayashi G, Lai OY, Dilthey A, Kuebler PJ, Wong TV, et al. Psoriasis patients are enriched for genetic variants that protect against HIV-1 disease. PLoS Genet. 2012;8:1–12.
112. Apps R, Qi Y, Carlson JM, Chen H, Gao X, Thomas R, et al. Influence of HLA-C expression level on HIV control. Science. 2013;340:87–91.
113. Kulkarni S, Qi Y, O'hUigin C, Pereyra F, Ramsuran V, McLaren P, et al. Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease. Proc Natl Acad Sci U S A. 2013;110:20705–10.
114. Chappell P, Meziane EK, Harrison M, Magiera L, Hermann C, Mears L, et al. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. Elife. 2015;2015:1–22.
115. Blander JM, Torchinsky MB, Campisi L. Revisiting the old link between infection and autoimmune disease with commensals and T helper 17 cells. Immunol Res. 2012;54:50–68.
116. Sfriso P, Ghirardello A, Botsios C, Tonon M, Zen M, Bassi N, et al. Infections and autoimmunity: the multifaceted relationship. J Leukoc Biol. 2010;87: 385–95.
117. Campisi L, Barbet G, Ding Y, Esplugues E, Flavell RA, Blander JM. Apoptosis in response to microbial infection induces autoreactive TH17 cells. Nat Immunol. 2016;17:1084–92.
118. Ercolini AM, Miller SD. The role of infections in autoimmune disease. Clin Exp Immunol. 2009;155:1–15.
119. Armstrong DL, Zidovetzki R, Alarcón-Riquelme ME, Tsao BP, Criswell LA, Kimberly RP, et al. GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region. Genes Immun. 2014;15:347–54.
120. McMurray RW, Elbourne K. Hepatitis C virus infection and autoimmunity. Semin Arthritis Rheum. 1997;26:689–701.
121. Maya R, Gershwin ME, Shoenfeld Y. Hepatitis B virus (HBV) and autoimmune disease. Clin Rev Allergy Immunol. 2008;34:85–102.
122. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol. 2002;2:647–55.
123. Baschal EE, Eisenbarth GS. Extreme genetic risk for type 1A diabetes in the post-genome era. J Autoimmun. 2008;31(1):1–6.
124. Zanelli E, Breedveld FC, De Vries RR. HLA class II association with rheumatoid arthritis: facts and interpretations. Hum Immunol. 2000;61:1254–61.
125. Rider LG. The heterogeneity of juvenile myositis. Autoimmun Rev. 2007;6(4): 241–7.
126. Libbey JE, Fujinami RS. Potential triggers of MS. Results Probl Cell Differ. 2010;51:21–42.
127. Smeekens SP, Ng A, Kumar V, Johnson MD, Plantinga TS, van Diemen C, et al. Functional genomics identifies type I interferon pathway as central for host defense against *Candida albicans*. Nat Commun. 2013;4:1342.
128. Gorer PA. The detection of a hereditary antigenic difference in the blood of mice by means of human group a serum. J Genet. 1936;32:17–31.
129. Bailey A, Dalchau N, Carter R, Emmott S, Phillips A, Werner JM, et al. Selector function of MHC I molecules is determined by protein plasticity. Sci Rep. 2015;5:14928.

# CHAPTER 6

## GENERAL DISCUSSION

Matzaraki V.

# CHAPTER 6

# GENERAL DISCUSSION

## • What We Learned From the Use of a Systems Genomics Approach in Candidaemia: From Genes to Potential Drug Targets

Traditionally, there are three major classes of antifungal drugs used for the treatment of *Candida* spp. bloodstream infections: 1) Triazoles (fluconazole, voriconazole), 2) echinocandins (caspofungin, micafungin, anidulafungin) and 3) polyenes (amphotericin B deoxycholate [AmB-d], liposomal amphotericin B [L-AmB], amphotericin B lipid complex [ABLC], and amphotericin B colloidal dispersion [ABCD]). The majority of these drugs target the cell wall (echinocandins) or the cell membrane (amphotericin, azoles) of the fungus[1–3]. Novel drug targets currently being investigated include glycosylphosphatidylinositol biosynthesis, histone deacetylases, and heat shock proteins[4]. However, it is important to mention that there is an alarming decrease in the development of antifungal drugs, with only five antifungal agents approved by the US Food and Drug Administration in the last 13 years[5]. One key fundamental challenge in the development of new antifungal drugs compared to other antimicrobials targeting bacteria is that fungal pathogens are more closely related to the human host, among other challenges that are discussed elsewhere[6]. Therefore, there is an urgent need to develop novel antifungal prophylactic or treatment strategies.

Given that individuals present considerable innate immune variation in host responses to candidaemia, ranging from extreme susceptibility to complete resistance and tolerance, host-directed therapies based on an individual's genetic make-up have a high probability of success. A recent study showed that drug mechanisms supported by genetic evidence would succeed twice as often as those without (from phase I to approval)[7]. In addition, a host-directed treatment may exert a weaker selection pressure on pathogens, potentially making it more difficult for a pathogen to evolve beyond the control of the host immune response. Thus, with the aim of selecting the best drug targets for designing more effective treatment strategies, our efforts should be directed at dissecting the genetic architecture of disease susceptibility and at understanding host-pathogen interactions in a more comprehensive way.

Given the complexity of host-pathogen interactions, conventional experimental approaches that study only individual molecular components (either of the host

or pathogen) cannot provide a comprehensive picture of these interactions. Thus, the integration of high-throughput multi-omics data (transcriptomics, proteomics, metabolomics, lipidomics, etc.), which can be used to prioritize genes for follow-up functional experiments to better understand their role in host immune defence and identify molecular pathways that underlie disease pathogenesis, is a necessity for advancing our understanding of host-pathogen interactions. These multi-omics data can be used to construct predictive models of the networks and dynamic interactions between the biological components that constitute the complex host-pathogen system. We followed this approach to prioritize the genes and identify the molecular pathways which underlie pathogenesis in *Candida* infections. Here we discuss directions that could be taken to develop new treatment strategies against candidaemia based on our findings (Fig.1).

In *chapter 2*, we demonstrated the potential of using a systems genomics approach to identify the genes and pathways that underlie susceptibility to candidaemia, ultimately prioritizing 31 genes enriched for the pathways inflammation, innate immunity, complement and hemostasis. Although the involvement of inflammation and innate immunity were expected findings, the identification of complement and hemostasis in candidaemia is novel. Given the vital role of the complement system in host immunity, and particularly in eliminating invading microorganisms, modulation of the complement system has been recognized as a promising strategy for drug discovery[8]. Thus, designing complement-specific drugs could be a promising strategy for candidaemia patients, and deserves further exploration.

A previously promising complement-specific drug target was one of our prioritized *cis*-genes, located in the candidaemia-associated locus, *PROC*. This gene was differentially expressed in response to *C. albicans* (Table 1, *chapter 2*). *PROC* encodes for protein C, which is known to reduce the clotting process, and its lack can lead to excess blood clotting[9,10]. This molecule had been previously targeted to treat sepsis, and the administration of recombinant activated protein C (APC) in sepsis patients had been suggested to increase the levels of protein C and ameliorate the patient outcome[11]. However, APC has also been associated with a higher risk of bleeding. In 2011, the European Medicines Agency withdrew Xigris® (human recombinant activated protein C), which was being produced by pharmaceutical company Eli Lilly, from the market and ongoing clinical trials were discontinued[11]. Although the use of APC was not successful for treating sepsis patients, our findings highlight the importance of blood coagulation processes during infection, particularly in candidaemia. However, it is worth keeping in mind that the biology of sepsis is complex, and not specific to infection, and success in developing new treatments will need stratification of patients based on genetic profiling so those with a high risk of sepsis will benefit most from the drug. Unless additional clinical trials provide evidence of a treatment effect based on genetic

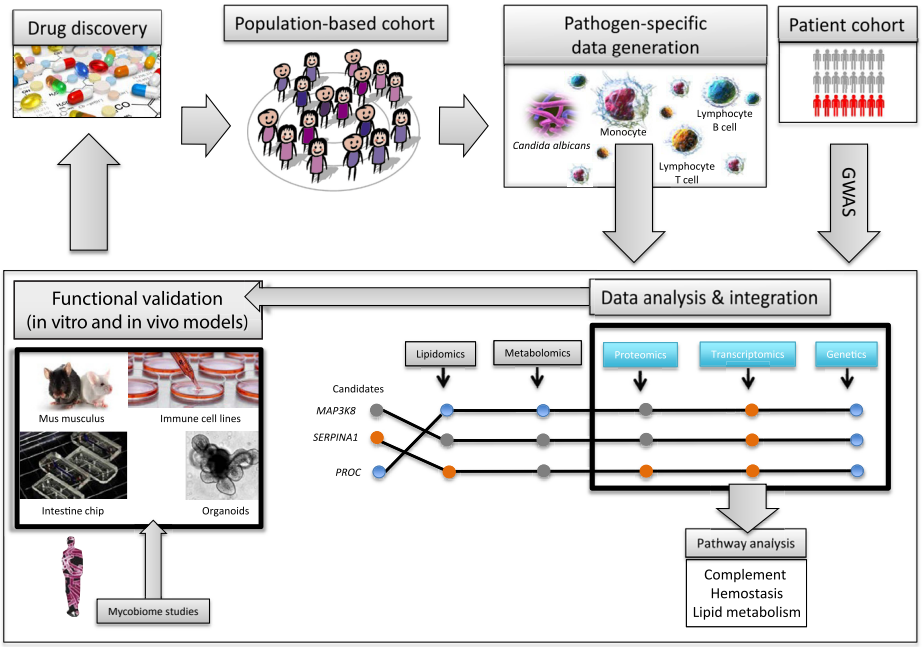



**Fig.1. 1: Summary of main findings on susceptibility to candidaemia and potential directions to be taken for discovery of new antifungal drugs.**
Given the limited size of patient cohorts, the use of a population-based cohort provides a promising platform to generate pathogen-specific molecular data. In this work, we used peripheral blood mononuclear cells (PBMCs) isolated from healthy volunteers (500FG cohort) stimulated with *C. albicans* to profile gene expression (using RNA sequencing, *chapter 2*), cytokines (using ELISA, *chapter 2* and *3*), and 92 circulatory inflammatory proteins (OLINK technology, *chapter 4*) in response to *C. albicans*. In addition to these molecular data, we performed the first GWAS on candidaemia using our candidaemia cohort consisting of 178 patients and 175 case-matched controls (*chapter 3*). Pathway enrichment analysis of differentially expressed genes in response to *C. albicans* revealed complement and hemostasis as important molecular pathways underlying pathogenesis of *Candida* infection, together with innate immunity and inflammation (*chapter 2*). Using genetic and cytokine measurements from the 500FG cohort, we mapped cytokine-QTLs in response to *C. albicans*. Of note, 21 independent loci that modulate cytokine levels were found to be associated with candidaemia susceptibility, and these were enriched for lipid metabolism pathways (*chapter 3*). By integrating different molecular data (blue labels: genetics, transcriptomics, proteomics), we prioritized genes for follow-up with *in vitro* and *in vivo* functional experiments. Three examples of genes discussed in this thesis are *MAP3K8*, *SERPINA1* and *PROC*. To obtain a more comprehensive picture of *C. albicans*-host interactions in the future, it is essential to make use of currently available high-throughput technologies (as summarized in Table 1, *chapter 1*) to generate additional molecular data (gray labels: metabolomics, lipidomics) in response to *C. albicans*. Integration of multi-omics data should identify interesting gene candidates for functional experiments using the most appropriate *in vitro* and/or *in vivo* models. Finally, mycobiome studies could provide additional valuable information about how the mycobiome and dysbiosis of gut health contribute to susceptibility to candidaemia. Ultimately, these multi-omics data can be used to construct predictive models of the networks and dynamic interactions that control candida infection and pinpoint promising drug targets for further investigation aimed at developing novel effective antifungal drugs.

profiling, we should be aware and cautious of the benefits and harms of APC, and in general of new drug targets.

Matzaraki V.

Early studies on developing protective antibodies (Ab) against experimental disseminated candidiasis in mice provided evidence that the mechanism of Ab protections is due to rapid and efficient complement deposition on the fungal cell surface, which leads to preferential uptake and killing by phagocytic cells[12–14]. In addition, mice with C5 deficiencies show decreased resistance in disseminated candidiasis[15,16]. To elucidate the role of the complement system in candidaemia, the use of animals with genetic deficiencies of individual complement components infected with *C. albicans* (such as C5-deficient mice and C6-deficient rats) would be particularly useful, as has been described for bacterial infections[8,17,18]. However, animal studies cannot completely represent the human situation. Thus, currently available proteomic technologies (mass spectrometry-based proteomics) are better suited to study the complement system to further understand fungal-host interactions. The recent development of a method of "mixed and quantitative proteome analysis", in which *C. albicans* and macrophages were simultaneously analyzed by nanoscale liquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS) without the need of isolation of the two individual living cells, provides new opportunities to study native interactions of *C. albicans* with macrophages (see Table 1, *Chapter 1*)[19]. NanoLC-MS/MS could be extended to additional immune cell types as regulation of the antifungal immune responses involves the coordinated contribution of different cell types.

Another interesting finding that needs further exploration as a potential therapeutic target is that we found that genetic variants associated with candidaemia susceptibility that influence cytokines upon *Candida* stimulation in peripheral blood mononuclear cells (PBMCs) are enriched for lipid and lipoprotein metabolism (*chapter 3*). This finding leads to the hypothesis that targeting host lipid synthesis and metabolism can potentially modulate cytokine responses and, subsequently, susceptibility to candidaemia. In general, inflammation and infections (by gram positive or negative bacterial, viral, tuberculosis) induce a variety of alterations in lipid metabolism that may initially dampen inflammation or fight infection, but which, if chronic, could contribute to the increased risk of atherosclerosis[20]. However, the exact mechanism by which inflammation and infection induce these lipid changes is still unclear. A hypothesis described by *P.S.R.K. Sastry* stated that the fungal infection could play a role in the development of atherosclerosis, which implies that *C. albicans* induces altered levels of lipids, but the mechanism by which this occurs is yet to be identified[21]. In line with this hypothesis, a recent study showed that fungi of the genus *Candida*, including *C. albicans*[22], were identified in 31.9% of atherosclerotic plaques. Based on these results, the authors speculated that *Candida* spp are capable of inducing inflammation of the vascular wall, and these in turn can lead to the development of atherosclerosis[22]. In addition, hyperlipoproteinemia enhances susceptibility to acute disseminated *C. albicans* infection in low-density-lipoprotein-receptor-deficient

(*Ldlr-/-*) mice in contrast to gram-negative bacterial infections[23]. In these studies, deficient *Ldlr* mice showed an increased outgrowth of *C. albicans* in their organs and produced significantly more pro-inflammatory cytokines than control mice[23]. For the future, quantitative proteomics coupled with traditional enzymatic colorimetric methods for follow-up confirmation of plasma lipids in patients would be useful for evaluating the host proteome responses upon *Candida* infection. Such a proteomic study recently revealed important changes in cellular structure and lipid metabolism in septic patients, changes that are possible targets for future intervention in sepsis[24].

In an attempt to further explore specific genes as potential drug targets, we validated the role of the *MAP3K8* (also known as *Tpl2*) gene in cytokine regulation in response to *Candida* stimulation (*chapter 2*). *MAP3K8* is located in one of the Immunochip-identified loci associated with candidaemia susceptibility, and encodes a member of the serine/threonine protein kinase family that has a critical role in innate immunity[25,26]. MAP3K8 signalling has been studied primarily downstream of TLR4 signaling and Tpl2(-/-) mice are susceptible to bacterial sepsis[26–28]. The association of the *MAP3K8* gene with susceptibility to *Candida* infection is novel and, in *chapter 2*, we demonstrated using chemical inhibitors that MAP3K8 modulates cytokine production in *Candida*-stimulated PBMCs. Another serine protease inhibitor that needs further exploration as a potential therapeutic target is α1-Antitrypsin (AAT), which is encoded by the gene *SERPINA1*. Several previous findings suggested that AAT has a potential protective role in host defence against various infectious diseases, including *Cryptosporidium parvum* infection, *Pseudomonas aeruginosa* lung infection in cystic fibrosis patients, and HIV-1 infection[29–32]. A recent proteomic study to identify both human serum proteins coating *C. albicans* cells and fungi surface proteins identified 13 proteins belonging to the serpin family, providing evidence that *SERPINA1* interacts with the host serum, and may therefore have a critical, as of yet unexplored, role in host immune defence upon *Candida* infection[33].

*MAP3K8* and *SERPINA1* thus seem to be interesting gene candidates to prioritize for future *in vitro* and *in vivo* experiments. Over-expressing or knocking-out these genes in relevant immune cell lines or mice, while challenging, would shed more light on their role in immune defence against *Candida* infection. Interestingly, the development of a murine model for *SERPINA1* has been challenging due to the presence of *SERPINA1* paralogs in the murine genome and the possibility of an embryonic role for *SERPINA1*, which results in embryonic lethality[34,35]. However, CRISPR/Cas9 technology has now enabled the creation of a new full knock out *SERPINA1* mouse model with a normal lifespan as a model for emphysema, and this model could also be highly relevant for studying the impact of *SERPINA1* in susceptibility to *Candida* infections[36].

In addition to working towards a new understanding of the genetics of susceptibility to candidaemia, we also investigated genetic variations that influence circulatory inflammatory proteins that could be used as potential biomarkers for *Candida* infection (*chapter 4*). Here we identified 10 novel genome-wide significant protein quantitative trait loci (pQTLs) that influence CCL4, MCP-1, VEGF-A, IL-8, MCP-2, MCP-3 and CXCL9 in *Candida*-stimulated PBMCs by performing genome-wide protein quantitative trail loci (pQTL) mapping. Intriguingly, none of the 10 genome-wide significant pQTLs strongly influence the expression of *cis*-genes, although most of them are intronic variants (or intergenic). However, we should keep in mind that there are multiple ways in which genetic variation can modulate the abundance of circulatory inflammatory proteins, such as regulation of translation and RNA editing, alternative splicing and protein release from the cell membrane. Of note, it has been shown that intronic SNPs within 200 base pairs of the splice junctions can have functional effects on splicing[37].

In addition to susceptibility, in *chapter 4*, we investigated the genetic variation that influences patient survival. Mortality rates in candidaemia patients range from 30-60%, suggesting that person-to-person differences, including genetics, may influence patient survival. We therefore performed a within-case GWAS to identify genetic variants that determine patient survival[38]. Of note, in contrast to our hypothesis that pQTLs explain susceptibility and/or patient survival, we showed that pQTLs showed a poor enrichment for genetic variants associated with either susceptibility to candidaemia or patient survival. This finding suggests that the genetic contribution to circulatory inflammatory responses in response to *C. albicans* is largely independent of the genetic contribution to disease susceptibility and patient survival. Moreover, we identified a genome-wide locus at chromosome 1 associated with survival ($P = 5x10^{-8}$) that did not show any significant association with susceptibility to candidaemia ($P = 0.097$). This is an interesting finding as it points to a different biology between disease susceptibility and patient survival that could provide new therapeutic opportunities for these two phenotypes.

To summarize, we followed a systems genomics approach to investigate genetics and molecular pathways that contribute to susceptibility to candidaemia (*chapter 2*). We also investigated which genetic variations contribute to the abundance of cytokines/circulatory inflammatory proteins upon *Candida* stimulation (*chapter 3* and *4*) and characterized variation that determines patient survival (*chapter 4*). Although there is still a long way to go to develop new antifungal drugs, we believe that the currently available technologies, together with advanced computational methods, promise exciting times in the field of systems genomics approaches to candidaemia.

## • Limitations, Challenges and Future Perspectives

Systems genomics approaches are providing new potentials to investigate genetics and molecular pathways underlying infectious diseases, including candidaemia. However there are still major limitations and challenges that apply not only to our studies on candidaemia but also to studies of other infectious diseases. First, all identified genetic associations should be validated in an independent cohort. However, this is not a possibility for most currently published associations for infectious diseases in the GWAS catalogue (https://www.ebi.ac.uk/gwas/). Independent cohorts are also not currently available for candidaemia, and the cohort we used is the largest candidaemia cohort available to date. In addition, GWAS on infectious diseases allows the identification of the most common polymorphisms, i.e. those showing a minor allele frequency (MAF) of at least ~5%.

In general, even in diseases where multiple loci have been identified by GWAS, the proportion of genetic variance that can be explained by all identified genetic associations is modest, often less than 20% of the total genetic variance estimated from twin studies[39]. This is because the individual associations identified by GWAS are generally of modest effect, with odds ratios of typically 1.1 to 1.5. This 'missing' heritability (the observation that loci detected by GWAS explain a small minority of the inferred genetic variance) may be explained by the presence of epistatic interactions (which we will discuss below) or structural variants, the search for which has met with little success so far[39].

An alternative hypothesis to explain the missing genetic component is the presence of rare variants, which may have larger effect sizes on disease phenotype but present low frequencies that cannot be detected with available GWAS microarrays. Previous GWAS studies from the National Human Genome Research Institute GWAS catalogue (https://www.ebi.ac.uk/gwas/) identified variants with a risk allele frequency <5%, indicating that low frequency and/or rare variants contribute to diverse phenotypes related to infectious diseases (Table 1). For candidaemia, three genome-wide significant, low frequency genetic variants ($P < 5x10^{-8}$) identified by Immunochip array were previously reported as associated to susceptibility[40], and we also reported several susceptibility variants of low frequency (between 1% and 5%) in *chapter 2* (Table 2). Thus, sequencing of just the exome of patients, composed of all the exons in the genome, would allow identification of these rare variants. Early studies using exon sequencing to identify rare variants in other infectious diseases were focused on *TLR4* gene in meningococcal disease and on five TLR genes on tuberculosis[41,42]. Both studies showed an excess of rare (and some more frequent) coding changes in patients compared to controls. Another recent study that discovered a susceptibility gene using exome sequenc-

ing and an extreme phenotype study design showed a strong association between presence of *DCTN4* missense variants (rs11954652, MAF=0.048, and rs35772018, MAF=0.0178) and susceptibility to both earlier and more severe *Pseudomonas aeruginosa* infection[43].

**Table 1.** Studies on infectious diseases from GWAS catalogue that identified low frequency variants (0.01 < MAF < 0.05) and rare variants (MAF < 0.01). The effect of variants was predicted using the Ensembl Variant Effect Predictor (VEP) online tool[70].

| Disease trait | Chr | pos | SNP | Effect of variant | Risk allele frequency | P value | OR/beta | Study |
|---|---|---|---|---|---|---|---|---|
| H1N1 severity | 4 | 2404594 | rs3135071 | intronic | 0.022 | $3x10^{-6}$ | 9.963 | Garcia-Etxebarria K[71] |
| HIV-1 control | 6 | 31464003 | rs2395029 | non-coding transcript exon variant | 0.032 | $1x10^{-25}$ | 5.3 | Pereyra F[72] |
| H1N1 infection | 6 | 130132737 | rs12528886 | intronic | 0.007 | $3x10^{-8}$ | 8.736 | Garcia-Etxebarria K[71] |
| H1N1 infection | 9 | 33218924 | rs28447319 | upstream gene variant | 0.004 | $1x10^{-14}$ | 21.08 | Garcia-Etxebarria K[71] |
| H1N1 infection | 9 | 28395334 | rs7851437 | intronic | 0.027 | $2x10^{-10}$ | 4.879 | Garcia-Etxebarria K[71] |
| H1N1 infection | 10 | 116427932 | rs41284366 | 5 prime UTR variant | 0.010 | $3x10^{-8}$ | 6.756 | Garcia-Etxebarria K[71] |
| H1N1 infection | 10 | 83884859 | rs10788289 | intergenic | 0.047 | $4x10^{-9}$ | 3.556 | Garcia-Etxebarria K[71] |
| H1N1 infection | 11 | 84624623 | rs17807909 | intronic | 0.011 | $5x10^{-15}$ | 10.09 | Garcia-Etxebarria K[71] |
| H1N1 infection | 12 | 70816898 | rs1040026 | intronic | 0.004 | $3x10^{-18}$ | 26.11 | Garcia-Etxebarria K[71] |
| H1N1 severity | 12 | 61554290 | rs11174072 | intergenic | 0.011 | $5x10^{-6}$ | 15.64 | Garcia-Etxebarria K[71] |
| H1N1 infection | 13 | 30921698 | rs41433250 | synonymous | 0.004 | $2x10^{-8}$ | 12.8 | Garcia-Etxebarria K[71] |
| Leprosy | 16 | 50820507 | rs16948876 | intergenic | 0.030 | $2x10^{-10}$ | 1.56 | Zhang F[73] |
| Tuberculosis | 16 | 75957548 | rs1948632 | intergenic | 0.040 | $8x10^{-6}$ | 1.55 | Thye T[74] |
| Tuberculosis | 16 | 82783986 | rs12386026 | intronic | 0.040 | $2x10^{-6}$ | 1.66 | Thye T[74] |
| Tuberculosis | 18 | 60441957 | rs4257308 | intergenic | 0.040 | $6x10^{-6}$ | 1.55 | Thye T[74] |
| H1N1 infection | X | 32217032 | rs16990169 | missense | 0.007 | $4x10^{-11}$ | 11.05 | Garcia-Etxebarria K[71] |
| *HIV-1 infection* | | | | | | | | |
| Setpoint viral load | 6 | | HLA-C*12:02 | | 0.010 | $4x10^{-8}$ | 0.478 | McLaren PJ[75] |
| Setpoint viral load | 6 | | HLA-B*52:01 | | 0.010 | $3x10^{-7}$ | 0.435 | McLaren PJ[75] |
| Setpoint viral load | 6 | | HLA-B*13:02 | | 0.030 | $4x10^{-10}$ | 0.346 | McLaren PJ[75] |
| Setpoint viral load | 6 | | HLA-B*14:02 | | 0.030 | $2x10^{-9}$ | 0.318 | McLaren PJ[75] |
| Setpoint viral load | 6 | | HLA-A*25:01 | | 0.030 | $7x10^{-8}$ | 0.301 | McLaren PJ[75] |
| Setpoint viral load | 6 | | HLA-B*27:05 | | 0.040 | $4x10^{-20}$ | 0.417 | McLaren PJ[75] |
| Setpoint viral load | 6 | | HLA-C*08:02 | | 0.040 | $9x10^{-11}$ | 0.307 | McLaren PJ[75] |
| Setpoint viral load | 6 | | HLA-C*01:02 | | 0.040 | $2x10^{-6}$ | 0.231 | McLaren PJ[75] |

Abbreviations: Chr, chromosome; pos, position (base pairs); OR, odds ratio; H1N1 severity, influenza A severity; H1N1 infection, influenza A infection

**Table 2.** Genetic variants that showed suggestive association with candidaemia susceptibility ($P < 9.99x10^{-5}$) (*chapter 2*) and those that reached genome-wide significance ($P < 5x10^{-8}$)[40], as identified using Immunochip array. The effect of variants was predicted using the Ensembl Variant Effect Predictor (VEP) online tool[70].

| Chr | SNP | Effect of variant | Frequency | OR | Study |
|---|---|---|---|---|---|
| 1 | rs3766122 | intronic | 0.04 | 2.02 | chapter 2 |
| 1 | rs296537 | intron variant, non-coding transcript variant | 0.01 | 3.67 | chapter 2 |
| 1 | rs17035850 | downstream gene variant | 0.03 | 4.68 | Kumar V[40] |
| 1 | rs4845320 | upstream gene variant | 0.01 | 4.25 | Kumar V[40] |
| 2 | rs6748999 | upstream gene variant | 0.04 | 2.21 | chapter 2 |
| 3 | rs12491812 | regulatory region variant | 0.01 | 3.24 | chapter 2 |
| 5 | rs16891982 | missense | 0.04 | 2.14 | chapter 2 |
| 6 | rs11760176 | non-coding transcript exon variant | 0.05 | 1.99 | chapter 2 |
| 6 | rs3127214 | intergenic | 0.03 | 2.96 | Kumar V[40] |
| 10 | rs1360119 | intronic | 0.01 | 3.62 | chapter 2 |
| 10 | rs7092540 | intergenic | 0.02 | 2.5 | chapter 2 |
| 14 | rs7149309 | intronic | 0.01 | 2.84 | chapter 2 |
| 16 | rs1802141 | downstream gene variant | 0.01 | 3.06 | chapter 2 |
| 17 | rs3848405 | intronic | 0.01 | 3.32 | chapter 2 |

Abbreviations: Chr, chromosome; OR, odds ratio

However, an argument can be made against focusing on exons because the majority of low frequency and/or rare variants that have been associated with the infectious diseases presented here, including candidaemia, are non-coding variants (intronic or intergenic) (Table 1 and 2). Interestingly, the majority of low frequency genetic variants associated with candidaemia susceptibility showed an intermediate effect, with an odds ratio (OR) between ~2 and ~5 (Table 2), as compared to low frequency variants associated to a number of immune-mediated diseases, which showed an OR between ~1.3 and ~4 (as extracted from GWAS catalogue, Table 3). In addition, infectious diseases such as influenza A (H1N1) infection, HIV-1 infection, tuberculosis and leprosy show a similar behavior to candidaemia, with low frequency variants showing an intermediate effect (or even as high as ~26 for influenza) (Table 1). Thus future studies of the genetics of infectious diseases, including candidaemia, should detect low frequency and/or rare variants using genomic tools (such as genotype imputation), custom genotyping arrays, and whole-genome sequencing in combination with the statistical approaches used for testing associations between phenotypes and rare variants that have made the study of these variants possible[44,45].

It is important to study the genetics of populations of different ethnicities, specifically those other than Europeans. This will provide both the opportunity to replicate already identified associations and a chance to identify new loci specific to the genetic background of a certain population. Moreover, the presence of gene-gene interactions (epistasis), which may mask the effect of certain genes on

**Table 3.** Studies on immune-mediated diseases from GWAS catalogue that identified low frequency variants (0.01 < MAF < 0.05). The effect of variants was predicted using the Ensembl Variant Effect Predictor (VEP) online tool[70].

| Disease trait | Chr | pos | SNP | Variant effect predictor | Risk allele frequency | P value | OR/beta | Study |
|---|---|---|---|---|---|---|---|---|
| SLE | 1 | 205540170 | rs181502228 | intergenic | 0.01 | $2x10^{-6}$ | 4.35 | Alarcon-Riquelme ME[76] |
| Atopic dermatitis | 1 | 152564174 | rs61813875 | downstream gene variant | 0.02 | $6x10^{-29}$ | 1.61 | Paternoster L[77] |
| SLE | 4 | 79049054 | rs10032909 | intronic | 0.01 | $1x10^{-6}$ | 2.42 | Alarcon-Riquelme ME[76] |
| Psoriasis | 6 | 31464003 | rs2395029 | non-coding transcript exon variant | 0.03 | $2x10^{-26}$ | 4.1 | Liu Y[78] |
| SLE | 6 | 137874586 | rs5029939 | intronic | 0.03 | $3x10^{-12}$ | 2.28 | Graham RR[79] |
| SLE | 6 | | DRB1*15:01 | | 0.05 | $7x10^{-12}$ | 2.1 | Alarcon-Riquelme ME[76] |
| SLE | 6 | | DQB1*06:02 | | 0.05 | $1x10^{-12}$ | 2.16 | Alarcon-Riquelme ME[76] |
| SLE | 6 | | DRB1*03:01 | | 0.05 | $4x10^{-11}$ | 1.94 | Alarcon-Riquelme ME[76] |
| SLE | 6 | | B*08:01 | | 0.03 | $1x10^{-8}$ | 2.06 | Alarcon-Riquelme ME[76] |
| Amyotrophic lateral sclerosis | 7 | 19257345 | rs10233425 | intergenic | 0.01 | $8x10^{-6}$ | 1.88 | Diekstra FP[80] |
| SLE | 9 | 128424504 | rs61732491 | missense | 0.01 | $2x10^{-6}$ | 3.36 | Alarcon-Riquelme ME[76] |
| IBD | 9 | 76302462 | rs11788518 | intronic | 0.03 | $7x10^{-6}$ | 1.39 | Yang SK[81] |
| SLE | 10 | 15496384 | rs7098187 | intergenic | 0.01 | $3x10^{-6}$ | 2.62 | Alarcon-Riquelme ME[76] |
| Crohn's disease | 12 | 40208138 | rs11175593 | intronic | 0.02 | $3x10^{-10}$ | 1.54 | Barrett JC[82] |
| IBD | 12 | 40398498 | rs11564258 | intronic | 0.03 | $6x10^{-29}$ | 1.334 | Jostins L[83] |
| Crohn's disease | 12 | 40398498 | rs11564258 | intronic | 0.03 | $6x10^{-21}$ | 1.74 | Franke A[84] |
| SLE | 15 | 48216203 | rs1878186 | intronic | 0.05 | $1x10^{-7}$ | 1.82 | Alarcon-Riquelme ME[76] |
| Crohn's disease | 16 | 50729868 | rs2066847 | frameshift | 0.02 | $3x10^{-24}$ | 3.99 | Barrett JC[82] |
| Crohn's disease | 16 | 67937477 | rs11574514 | downstream gene variant | 0.05 | $2x10^{-7}$ | 1.44 | Kenny EE[85] |

Abbreviations: Chr, chromosome; pos, position (base pairs); OR, odds ratio; SLE, systemic lupus erythematosus; IBD, inflammatory bowel disease

disease outcome, makes it challenging to study the genetics of infectious diseases. This means it is also important to study epistatic interactions. Of note, a previous study showed that epistatic interactions in fruit flies greatly influence the variance in quantitative traits, or characteristics influenced by multiple genes[46]. Therefore, gaining insight into epistatic effects can be used to infer genetic networks affecting quantitative traits. Knowing how genes interact with each other, and the exact effect of a variant, is valuable information to advance discoveries in personalized medicine. We reviewed epistatic interactions in autoimmune and infectious diseases with a focus on the MHC locus in *chapter 5*.

Most of our molecular data have been produced in a mixture of immune cells, such as PBMCs, (*chapters 2*, *3* and *4*) because they are an easily accessible cell system. However, PBMCs are a mixture of immune cells consisting of T cells, B cells, NK cells, and monocytes and they do not contain neutrophils, erythrocytes and platelets. Therefore, using only PBMCs may not reflect the full range of physiological responses in human. Additionally, the use of animal models could provide a useful setting to test the role of gene candidates in immune defence against *C. albicans*. Animal models have been widely used to study host-pathogen interactions in an attempt to mirror the pathophysiology of most diseases, but may not always faithfully mimic human disease due to inter-species variations[47]. Instead, to fill the gap between standard 2D cell culture and animal models, organoids have been widely used to model host-pathogen interactions[48]. One of the unique advantages of organoids is that they allow researchers to examine patient-specific responses to pathogens as they are derived from two sources: (1) pluripotent stem cells (PSCs) or (2) adult stem cells (ASCs) derived from patients[48]. However, organoids are heterogeneous in both shape and size, which may lead to inconsistent findings. One potentially innovative way to overcome such challenges is microengineered models of the functional units of human organs, known as 'organs-on-chips'[49]. 'Organ-on-chips' technology offers the possibility to emulate kidney, liver, brain, lung, intestines, and heart on a chip while also simulating the micro-environment that cells actually need to behave as if they are in the human body. Thus, these mini-organs provide a good alternative to using animal models.

Given that an effective immune response depends on the proper activation, regulation, function, and interaction of immune cells with their micro-environment (which includes a range of stimuli and other cell types such as endothelial cells[50]), 'organ-on-chips' would allow the study of these interactions in a more dynamic environment that is comparable to human physiology. An artificial liver was recently developed using this technology to test its responses to hepatitis B virus (HBV) infection[51]. This liver-on-a-chip had similar biological responses, including innate immune and cytokine responses, to those seen in the human liver following infection with HBV, a demonstration of the potential of using this new

technology for studying host-pathogen interactions. For *Candida* infection, it has been observed that fungal burden changes with variable dynamics in the kidney, brain, spleen, and liver, and declines in all organs except for the kidney in a fatal mouse model of invasive candidiasis[52]. Thus it would be interesting to study organ-specific control of *Candida* using organ-on-a-chip technology to identify which cellular and molecular factors determine these differences between organs. Moreover, in the long run, these organs on chips will revolutionize the development of drugs in personalized medicine as custom chips (potentially created by using stem cells from an actual patient) that imitate a patient's unique biology would allow us to better predict how that patient will respond to specific drugs used to treat their infection[53].

Another important aspect to consider in studying host-pathogen interactions is that we need to understand in which cell types the changes at transcriptomic, proteomic or metabolomic level occur upon infection. For this, single-cell methods of analyzing the transcriptome, proteome and metabolome will provide higher resolution that can pinpoint which cell types are the major players during infectious disease. Several immunological studies have already demonstrated the role of monocytes and neutrophils in clearing *Candida* infection[54,55], however single-cell RNA sequencing may reveal new context-specific *Candida*-specific signatures and novel cell types that are involved in the complex immune response against the fungal pathogens. To identify regulatory networks that govern immune response during infection, ChiP-Seq, ATAQseq and PRO-seq are the key technologies currently available (Table 1, *chapter 1*) to build these regulatory networks. Integration of molecular data generated by these technologies with results from assays that interrogate the status of DNA, RNA, proteins, lipids and metabolites should provide a move comprehensive picture of host-pathogen interactions. In the near future, we should direct our efforts at generating all these multiple layers of molecular information, which can then be used to construct predictive models for the development of specific therapeutics.

Another challenge in using a systems genomics approach in infectious diseases is the development of computational methods robust enough to handle and analyze all these large -omics data and building relevant models to predict disease outcome and/or drug targets given the complexity of the host-pathogen interactions. Of note, a systems genomics approach makes use of large different datasets which requires cooperation and communication among the diverse scientists involved in experimental design, sample preparation, high-throughput assays and analysis and integration of large datasets. Therefore, building bridges between diverse members of the scientific community is crucial to advancing our knowledge using a systems approach. Another urgent challenge is that we need to establish publically available databases to store and share the ever-increasing pathogen-specific

-omics datasets. The 'Systems Biology Program for infectious Disease Research' consortium funded by the US National Institute of Health (NIH) represents an example of this direction; it provides experimental datasets, software tools, computational models, and research protocols through the four NIH centers' and the National Institute of Allergy and Infectious Diseases websites (http://www.niaid.nih.gov/labsandresources/resources/sb/Pages /default.aspx)[56].

We also need to raise awareness about the need to increase the numbers of people included in patient cohorts, which are the main sources for generating -omics datasets to answer our research questions. For this, the scientific community first needs to clearly communicate with patients the importance of this work and to share all the information needed to help participants make an informed decision about participation. Next we need a coordinated effort aimed at collecting samples from a hospital-based setting that feeds into an appropriate infrastructure for archiving all biological samples from donors for use in research (biobanking) and to define ways of ensuring the sustainability of biobanking. However, while collecting cohorts and samples may be feasible for infectious diseases with millions of patients such as HIV (the World Health Organization estimated there were 36.7 million people living with HIV at the end of 2016), this will be more challenging for relatively rare diseases like invasive *Candida* infections. This difference is reflected in the number of currently available cohorts for HIV-related research, which the Harvard University Center for AIDS Research (https://cfar.globalhealth.harvard.edu/) currently catalogues as 30, a number that is in sharp contrast to our candidaemia cohort which is both unique and the largest to date.

Further specific limitations in our studies include that fact that our transcriptomic experiments on *Candida*-stimulated PBMCs were performed at only two time points, 4 and 24 hours (*chapter 2*) even though gene transcription is a dynamic process through time. Thus we ideally need to capture gene transcription at more time points to capture early and late *Candida*-specific gene signatures. A recent transcriptomic study of neutrophils, the main effector cells in killing the fungus, stimulated with both morphotypes of *C. albicans*, showed that the strongest transcriptional responses was after 60 min, independent of the *Candida* morphotype[57]. In this study, the infection of neutrophils was studied after 15, 30 and 60 mins to represent different stages of *in vitro* infection: recognition and adhesion (15 min), early phagocytosis (30 min), and entire phagocytosis of yeast or wrapping of hyphae (60 min)[57]. Thus, as discussed earlier, we need to extend the transcriptomic studies to a single-cell level by using multiple time points that reflect different stages of infection to achieve higher resolution than using only PBMCs. Moreover, we performed our studies using heat-killed *C. albicans* yeast, which is fundamentally different from live *C. albicans*. In PBMC cultures, live *C. albicans* can suppress IL-17 production during anti-*Candida* host immune de-

fence[58,59]. This makes it difficult to delineate defence mechanisms upon live *C. albicans* stimulation, rendering heat-killing an unavoidable necessity in our studies. Another focus for future studies should be testing other medically relevant fungal species than *C. albicans*, such as *C. tropicalis*, *C. glabrata*, and *C. krusei*[60].

In addition, given that *C. albicans* is part of the normal microbiome composition in the host's gastrointestinal tract, fungal microbiome (also known as the mycobiome) studies would provide an additional valuable layer of information for better understanding how the mycobiome contributes to candidaemia susceptibility. For instance, in a recent study using a murine oral candidiasis model, mice treated with *Pichia* spent medium showed significantly lower infection scores, fungal burden, and tongue epithelial damage, indicating these findings from mycobiome data can lead to potential novel antifungals[61]. Further questions that could be addressed in mycobiome studies would be how the mycobiome influences the immune system and whether genetic factors contribute to mycobiome composition. Studying the mycobiome, and particularly *C. albicans* within it, may lead to the discovery of new metabolic pathways and key steps that play an important function in the pathogen proliferation, hyphae formation and pathogenesis that are either absent in human or different enough from their human counterparts. These distinct pathways from the human host could be potential targets for the development of novel drugs that show a reduced toxicity in the host.

Moreover, understanding host-pathogen interactions could shed light on the potential role of infectious agents as triggers for autoimmune diseases (a hypothesis discussed in detail in *chapter 5*). It would be interesting to study whether an opportunistic infection, such as *Candida* infection, can contribute to the development of autoimmune diseases. A recent study, for example, found that translocation of a gut microbe, *Enterococcus gallinarum*, from the gut to the liver and other systemic tissues triggers autoimmune responses in a mouse predisposed to autoimmunity[62]. Another study showed a possible mechanism for how infections alter genetic risk to make some individuals more susceptible to inflammatory diseases. The study showed that a protein produced by the Epstein-Barr virus (EBNA2) interacts with nearly half of the genetic risk loci associated with lupus in Europeans[63]. This interaction may contribute to changes in gene expression that in turn increase an individual's chance of developing lupus[63]. Similar EBNA2-anchored associations exist in multiple sclerosis, rheumatoid arthritis, inflammatory bowel disease, Type 1 diabetes, juvenile idiopathic arthritis and celiac disease[63]. Lastly, it has been hypothesized that *C. albicans* may trigger celiac disease because the yeast protein Hwp1 presents sequence homology with the gluten protein gliadin, which is a known trigger for celiac disease[64]. Recent studies have observed humoral cross reactivity between Hwp1 and gliadin during celiac disease and *C. albicans* infection, which supports the hypothesis that *Candida* infection triggers onset of celiac

disease in genetically-susceptible individuals[65].

With the aim of establishing large-scale host-fungi interactions networks on a systems-biology scale, an extended inference approach based on protein orthology and data on gene function was recently developed for *C. albicans* and *Aspergillus fumigatus* (*A. fumigatus*)[66]. This integrated network framework can serve as a basis for analyses of high-throughput host-fungi transcriptome and proteome data[66]. This newly developed framework highlights that there is an awareness of directing our efforts to generate and integrate multi-omics data to better understand host-fungi interactions, which is underestimated in literature compared to bacterial and viral interactions. To sum up, given the inherent complexity of the human immune response to encountering an invading pathogen such as *C. albicans*, a systems genomics approach seems appropriate and promising for furthering our knowledge of pathogen-host interactions. Ultimately, this knowledge will help us develop more effective strategies that modulate immune responses that can help patients better control and fight infections.

## • Concluding Remarks

Our studies allowed us a glimpse of the genetic architecture of candidaemia. We observed that the majority of low frequency genetic variants associated with candidaemia susceptibility (with a frequency between 1% and 5%) showed an intermediate effect (Fig. 2, Table 2). From an evolutionary perspective, we can assume that variants that increase disease susceptibility are deleterious to fitness, and should be selected against, with the consequence that susceptibility variants should be low frequency and/or rare[67]. In addition, considering that mutation rates are sufficiently high, selection cannot remove all deleterious variants, and those that have an intermediate/modest effect on fitness may reach allele frequencies of 1% or occasionally more[67]. Moreover, given that *Candida* infections are opportunistic infections, we can speculate that these infections have co-evolved with other more prevalent infectious diseases, complicating the process of how the architecture of the human genome has been shaped by opportunistic infections such as *Candida* infections. For instance, during the 1980's, the emergence of the highly variable and fast evolving retroviral pathogen, HIV-1, that progressively causes a failure of the immune system would potentially have allowed opportunistic infections, such as fungal infections, to take hold. Indeed, the rise of human fungal infections coincided largely with the onset of the AIDS pandemic of the 1980's (https://ensia.com/features/19036/). Of note, candidaemia appears to be common in HIV-infected patients, particularly in patients with advanced HIV disease[68].

Overall, we can speculate that three theories can explain the genetic architecture

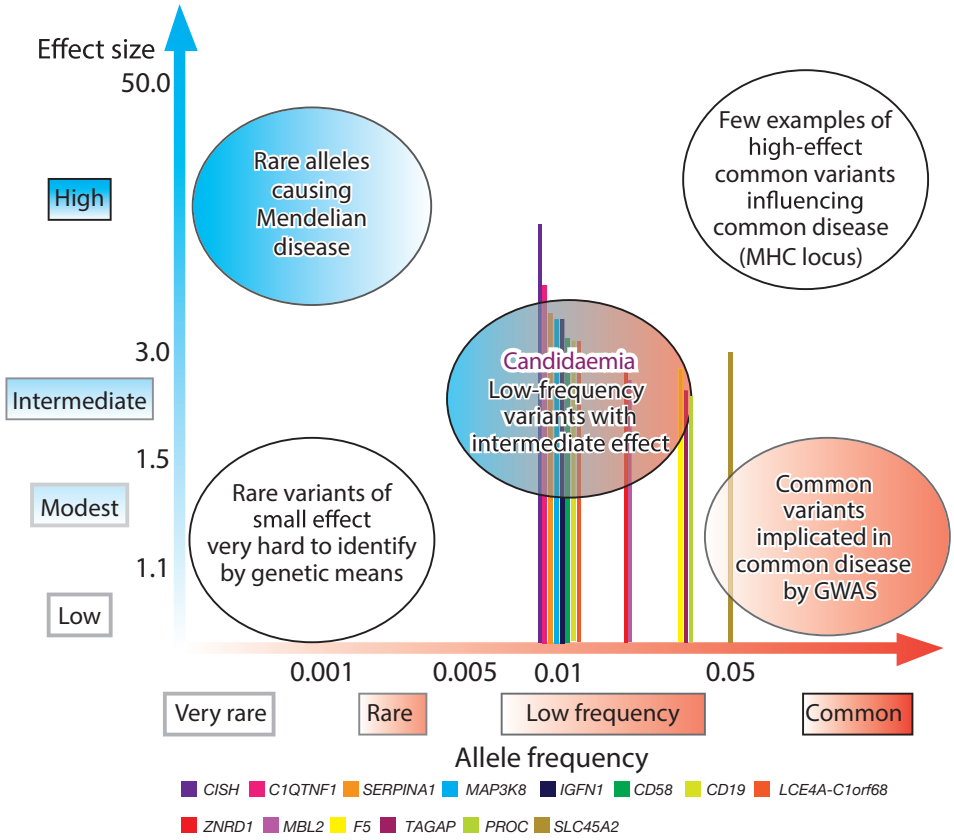



**Fig.2. Allele frequency and effect sizes for genetic variants associated with candidaemia susceptibility.** Solid colored bars (not depicted to scale) show the allele frequencies and Odds Ratios (ORs) of genetic variants associated with candidaemia susceptibility as presented in Table 2. The colored labels at the bottom correspond to *cis*-genes in close proximity to susceptibility variants (in a window of 500 kb). Figure adapted from that of *Manolio T.A. et al*[39].

of *Candida* infections. Theory 1 suggests that relatively common variants account for susceptibility. If common variants are of importance to susceptibility, an extension of our GWAS approach of analysis of larger sample sizes might be the best way to identify common variants that confer susceptibility to the disease. In contrast, theory 2 speculates that very rare mutations account for susceptibility, as in primary immunodeficiency diseases. Indeed, several primary immunodeficiencies that behave as monogenetic disorders were found to show an increased susceptibility to *Candida* infections[69]. Lastly, theory 3 suggests that many individually rare variants with incomplete penetrance account for the genetic component of candidaemia susceptibility. If rare variants account for most susceptibility, next-generation sequencing of ever larger human cohorts should allow identifica-

tion of rare variants that can explain susceptibility.

In summary, our studies have opened new opportunities to understand the genetic architecture of candidaemia susceptibility and to identify genetic variants, genes and pathways that are implicated in disease susceptibility. Importantly, we have provided preliminary evidence that systems genomics approaches like ours can be a useful tool for identifying novel risk genes and pathways for *Candida* infection, especially given that collection of large cohorts is difficult because invasive *Candida* infections are relatively scarce. Despite the challenges it presents, we acknowledge that a systematic integration of different layers of molecular data may be the only avenue to further our knowledge and understanding of the complex host-pathogen interactions that shape our human genome. Last, but not least, this ever-growing body of molecular data will play a critical role in the development of novel prophylactic and treatment strategies that target relevant genes and pathways to control and limit infectious diseases.

# 10 Take-home messages

1. The genetic architecture of infectious diseases, including candidaemia, seems to differ from that of non-infectious diseases (such as immune-mediated diseases), with low frequency variants showing an intermediate effect in infectious diseases (Fig. 2).

2. A systems genomics approach has the potential to obtain novel insights into the genetics and molecular mechanisms underlying disease pathogenesis in relatively rare infections such as candidaemia (*chapter 2*).

3. Hemostasis seems to be an important biological process involved in susceptibility to candidaemia (*chapter 2*).

4. Host cytokine responses, as induced by *C. albicans*, *A. fumigatus* and *C. neoformans*, are organized in a pathogen-specific manner (*chapter 3*).

5. Cytokine responses induced by *C. albicans* yeast and hyphae are very similar, suggesting that cytokine responses to these two morphotypes are regulated by similar genetic and non-genetic factors (*chapter 3*).

6. Genes located in loci associated with susceptibility to candidaemia that modulate cytokine levels upon *C. albicans* stimulation are enriched in lipid

and lipoprotein metabolic processes (*chapter 3*).

**7.** Our first GWAS on candidaemia identified a novel strong genetic association of candidaemia with polymorphisms in the *PLA2G4B* gene that encodes for the cytosolic phospholipase A2 protein (*chapter 3*).

**8.** By comparing SNPs and alleles in the MHC locus associated with autoimmune and infectious diseases, we observed alleles with opposite allelic effect, supporting the 'pathogen-driven selection' hypothesis, and alleles with common genetic signatures, indirectly supporting the second hypothesis that pathogens can trigger autoimmunity (*chapter 5*).

**9.** Genetic variants influencing levels of circulatory inflammatory levels (protein-QTLs) showed a poor enrichment for variants associated with candidaemia susceptibility and patient survival, implying a distinct genetic contribution between inflammatory responses, candidaemia susceptibility and patient survival (*chapter 6*).

**10.** Follow-up functional studies are essential to shed light on the role of genes and molecular mechanisms in host defence against *Candida* infection to better understand their potential as therapeutic targets.

# References

1. Tada, R., Latgé, J. P. & Aimanianda, V. Undressing the fungal cell wall/cell membrane--the antifungal drug targets. Curr. Pharm. Des. 19, 3738 (2013).
2. Georgopapadakou, N. H. & Walsh, T. J. Human mycoses: drugs and targets for emerging pathogens. Science (80-. ). 264, 371 LP-373 (1994).
3. Ghannoum, M. A. & Rice, L. B. Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. Clin.Microbiol.Rev. 12, 501–517 (1999).
4. McCarthy, M. W., Kontoyiannis, D. P., Cornely, O. A., Perfect, J. R. & Walsh, T. J. Novel agents and drug targets to meet the challenges of resistant fungi. The Journal of infectious diseases 216, S474–S483 (2017).
5. Ostrosky-Zeichner, L., Casadevall, A., Galgiani, J. N., Odds, F. C. & Rex, J. H. An insight into the antifungal pipeline: Selected new molecules and beyond. Nature Reviews Drug Discovery 9, 719–727 (2010).
6. Roemer, T. & Krysan, D. J. Antifungal drug development: challenges, unmet clinical needs, and new approaches. Cold Spring Harbor perspectives in medicine 4, (2014).
7. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. Nat. Genet. 47, 856–860 (2015).
8. Kirschfink, M. & Mollnes, T. E. Modern complement analysis. Clin. Diagn. Lab. Immunol. 10, 982–989 (2003).
9. Esmon, C. T. The protein C pathway. Chest 124, (2003).

10. Dahlbäck, B. & Villoutreix, B. O. The anticoagulant protein C pathway. FEBS Letters 579, 3310–3316 (2005).
11. Martí-Carvajal, A. J., Solà, I., Gluud, C., Lathyris, D. & Cardona, A. F. Human recombinant protein C for severe sepsis and septic shock in adult and paediatric patients. Cochrane database Syst. Rev. 12, CD004388 (2012).
12. Han, Y. & Cutler, J. E. Assessment of a mouse model of neutropenia and the effect of an anti-candidiasis monoclonal antibody in these animals. J. Infect. Dis. 175, 1169–75 (1997).
13. Germann, T. et al. Interleukin-12 profoundly up-regulates the synthesis of antigen-specific complement-fixing IgG2a, IgG2b and IgG3 antibody subclasses in vivo. Eur. J. Immunol. 25, 823–829 (1995).
14. Qian, Q., Jutila, M. A., Van Rooijen, N. & Cutler, J. E. Elimination of mouse splenic macrophages correlates with increased susceptibility to experimental disseminated candidiasis. J. Immunol. 152, 5000–8 (1994).
15. Sohnle, P. G., Frank, M. M. & Kirkpatrick, C. H. Deposition of complement components in the cutaneous lesions of chronic mucocutaneous candidiasis. Clin. Immunol. Immunopathol. 5, 340–350 (1976).
16. Hector, R. F., Domer, J. E. & Carrow, E. W. Immune responses to Candida albicans in genetically distinct mice. Infect. Immun. 38, 1020–1028 (1982).
17. Rother K, R. U. Hereditary and acquired complement deficiencies in animals and man. Prog Allergy 39, 1–7 (1986).
18. Mold, C. Role of complement in host defense against bacterial infection. Microbes Infect. 1, 633–638 (1999).
19. Kitahara, N. et al. Description of the interaction between Candida albicans and macrophages by mixed and quantitative proteome analysis without isolation. AMB Express 5, (2015).
20. Kenneth R Feingold, and Carl Grunfeld. The Effect of Inflammation and Infection on Lipids and Lipoproteins. (South Dartmouth (MA), 2015). at <https://www.ncbi.nlm.nih.gov/books/NBK326741/>
21. Sastry, P. S. R. K. Occult fungal infection is the underlying pathogenic cause of atherogenesis. Medical Hypotheses 63, 671–674 (2004).
22. Nurgeldiyeva, M. J., Hojakuliyev, B. G. & Muhammedov, M. B. Correlation of atherogenesis with an infection of Candida albicans. Int. J. Clin. Exp. Med. 7, 2137–2143 (2014).
23. Netea, M. G. et al. Hyperlipoproteinemia enhances susceptibility to acute disseminated Candida albicans infection in low-density-lipoprotein-receptor-deficient mice. Infect. Immun. 65, 2663–2667 (1997).
24. Sharma, N. K. et al. Proteomic study revealed cellular assembly and lipid metabolism dysregulation in sepsis secondary to community-acquired pneumonia. Sci. Rep. 7, 15606 (2017).
25. Patriotis, C., Makris, A., Bear, S. E. & Tsichlis, P. N. Tumor progression locus 2 (Tpl-2) encodes a protein kinase involved in the progression of rodent T-cell lymphomas and in T-cell activation. Proc. Natl. Acad. Sci. 90, 2251–2255 (1993).
26. Mielke, L. a et al. Tumor progression locus 2 (Map3k8) is critical for host defense against Listeria monocytogenes and IL-1 beta production. J. Immunol. 183, 7984–7993 (2009).
27. Mieulet, V. et al. TPL-2-mediated activation of MAPK downstream of TLR4 signaling is coupled to arginine availability. Sci. Signal. 3, (2010).
28. Banerjee, A., Gugasyan, R., McMahon, M. & Gerondakis, S. Diverse Toll-like receptors utilize Tpl2 to activate extracellular signal-regulated kinase (ERK) in hemopoietic cells. Proc. Natl. Acad. Sci. 103, 3274–3279 (2006).
29. Forney, J. R., Yang, S. G. & Healey, M. C. Interaction of the human serine protease inhibitor alpha-1-antitrypsin with Cryptosporidium parvum. J. Parasitol. 82, 496–502 (1996).
30. Griese, M. et al. alpha1-Antitrypsin inhalation reduces airway inflammation in cystic fibrosis patients. Eur. Respir. J. 29, 240–50 (2007).
31. Münch, J. et al. Discovery and Optimization of a Natural HIV-1 Entry Inhibitor Targeting the gp41 Fusion Peptide. Cell 129, 263–275 (2007).
32. Forssmann, W. G. et al. Short-term monotherapy in HIV-infected patients with a virus entry inhibitor against the gp41 fusion peptide. Sci. Transl. Med. 2, (2010).
33. Marín, E. et al. Candida albicans shaving

to profile human serum proteins on hyphal surface. Front. Microbiol. 6, (2015).
34. Barbour, K. W. et al. The murine α1-proteinase inhibitor gene family: Polymorphism, chromosomal location, and structure. Genomics 80, 515–522 (2002).
35. Wang, D. et al. Deletion of serpina1a, a murine α1-antitrypsin ortholog, results in embryonic lethality. Exp. Lung Res. 37, 291–300 (2011).
36. Borel, F. et al. Editing out five Serpina1 paralogs to create a mouse model of genetic emphysema. Proc. Natl. Acad. Sci. U. S. A. 115, 2788–2793 (2018).
37. Lalonde, E. et al. RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. Genome Res. 21, 545–554 (2011).
38. Flevari A., Theodorakopoulou M., Velegraki A., Armaganidis A., D. G. Treatment of invasive candidiasis in the elderly: a review. Clin Interv Aging 8, 1199–1208 (2013).
39. Manolio, T. A. et al. Finding the missing heritability of complex diseases. Nature 461, 747–53 (2009).
40. Kumar, V. et al. Immunochip SNP array identifies novel genetic variants conferring susceptibility to candidaemia. Nat. Commun. 5, 4675 (2014).
41. Smirnova, I. et al. Assay of locus-specific genetic load implicates rare Toll-like receptor 4 mutations in meningococcal susceptibility. Proc. Natl. Acad. Sci. 100, 6075–6080 (2003).
42. Ma, X. et al. Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PLoS One 2, (2007).
43. Emond, M. J. et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat. Genet. 44, 886–889 (2012).
44. Feng, T. & Zhu, X. Detecting rare variants. Methods Mol. Biol. 850, 453–464 (2012).
45. Bomba, L., Walter, K. & Soranzo, N. The impact of rare and low-frequency genetic variants in common disease. Genome Biology 18, (2017).
46. Huang, W. et al. Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proc. Natl. Acad. Sci. 109, 15553–15559 (2012).
47. Conti, F., Abnave, P. & Ghigo, E. Unconventional animal models: a booster for new advances in host—pathogen interactions. Front. Cell. Infect. Microbiol. 4, 142 (2014).
48. Bartfeld, S. Modeling infectious diseases and host-microbe interactions in gastrointestinal organoids. Developmental Biology 420, 262–270 (2016).
49. Esch, E. W., Bahinski, A. & Huh, D. Organs-on-chips at the frontiers of drug discovery. Nature Reviews Drug Discovery 14, 248–260 (2015).
50. Danese, S., Dejana, E. & Fiocchi, C. Immune Regulation by Microvascular Endothelial Cells: Directing Innate and Adaptive Immunity, Coagulation, and Inflammation. J. Immunol. 178, 6017–6022 (2007).
51. Ortega-Prieto, A. M. et al. 3D microfluidic liver cultures as a physiological preclinical tool for hepatitis B virus infection. Nat. Commun. 9, (2018).
52. Romani, L. et al. An immunoregulatory role for neutrophils in CD4+ T helper subset selection in mice with candidiasis. J. Immunol. 158, 2356–62 (1997).
53. Workman, M. J. et al. Enhanced Utilization of Induced Pluripotent Stem Cell–Derived Human Intestinal Organoids Using Microengineered Chips. CMGH (2018). doi:10.1016/j.jcmgh.2017.12.008
54. Fradin, C. et al. Granulocytes govern the transcriptional response, morphology and proliferation of Candida albicans in human blood. Mol. Microbiol. 56, 397–415 (2005).
55. Netea, M. G. et al. Human dendritic cells are less potent at killing Candida albicans than both monocytes and macrophages. Microbes Infect 6, 985–989 (2004).
56. Aderem, A. et al. A systems biology approach to infectious disease research: Innovating the pathogen-host research paradigm. MBio 2, (2011).
57. Niemiec, M. J. et al. Dual transcriptome of the immediate neutrophil and Candida albicans interplay. BMC Genomics 18, 696 (2017).
58. Cheng, S. C. et al. Candida albicans dampens host defense by downregulating IL-17 production. J. Immunol. 185, 2450–2457 (2010).
59. van de Veerdonk, F. L. et al. The macrophage mannose receptor induces IL-17 in re-

sponse to Candida albicans. Cell Host Microbe 5, 329–340 (2009).
60. Campion, E. W., Kullberg, B. J. & Arendrup, M. C. Invasive Candidiasis. N. Engl. J. Med. 373, 1445–1456 (2015).
61. Mukherjee, P. K. et al. Oral mycobiome analysis of HIV-infected patients: identification of pichia as an antagonist of opportunistic fungi. PLoS Pathog. 10, (2014).
62. Manfredo Vieira, S. et al. Translocation of a gut pathobiont drives autoimmunity in mice and humans. Science (80-. ). 359, 1156–1161 (2018).
63. Harley, J. B. et al. Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. Nat. Genet. (2018). doi:10.1038/s41588-018-0102-3
64. Nieuwenhuizen, W. F., Pieters, R. H. H., Knippels, L. M. J., Jansen, M. C. J. F. & Koppelman, S. J. Is Candida albicans a trigger in the onset of coeliac disease? Lancet 361, 2152–2154 (2003).
65. Corouge, M. et al. Humoral immunity links Candida albicans infection and celiac disease. PLoS One 10, (2015).
66. Remmele, C. W. et al. Integrated inference and evaluation of host-fungi interaction networks. Front. Microbiol. 6, (2015).
67. Gibson, G. Rare and common variants: Twenty arguments. Nature Reviews Genetics 13, 135–145 (2012).
68. Tumbarello, M. et al. Candidemia in HIV-infected subjects. Eur. J. Clin. Microbiol. Infect. Dis. 18, 478–483 (1999).
69. Smeekens, S. P., van de Veerdonk, F. L., Kullberg, B. J. & Netea, M. G. Genetic susceptibility to Candida infections. EMBO Mol. Med. 5, 805–813 (2013).
70. McLaren, W. et al. The Ensembl Variant Effect Predictor. Genome Biol. 17, (2016).
71. Garcia-Etxebarria, K. et al. No major host genetic risk factor contributed to A(H1N1)2009 influenza severity. PLoS One 10, e0141661 (2015).
72. International HIV Controllers Study, T. I. H. C. et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330, 1551–7 (2010).
73. Zhang, F. et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. Nat. Genet. 43, 1247–1251 (2011).
74. Thye, T. et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat. Genet. 42, 739–741 (2010).
75. McLaren, P. J. et al. Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. Proc. Natl. Acad. Sci. U. S. A. 1–6 (2015). doi:10.1073/pnas.1514867112
76. Alarcón-Riquelme, M. E. et al. Genome-Wide Association Study in an Amerindian Ancestry Population Reveals Novel Systemic Lupus Erythematosus Risk Loci and the Role of European Admixture. Arthritis Rheumatol. 68, 932–943 (2016).
77. Paternoster, L. et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat. Genet. 47, 1449–1456 (2015).
78. Liu, Y. et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS Genet. 4, (2008).
79. Graham, R. R. et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat. Genet. 40, 1059–1061 (2008).
80. Diekstra, F. P. et al. C9orf72 and UNC13A are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: A genome-wide meta-analysis. Ann. Neurol. 76, 120–133 (2014).
81. Yang, S. K. et al. Identification of Loci at 1q21 and 16q23 That Affect Susceptibility to Inflammatory Bowel Disease in Koreans. Gastroenterology 151, 1096–1099.e4 (2016).
82. Barrett, J. C. et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat. Genet. 40, 955–962 (2008).
83. Jostins, L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491, 119–124 (2012).
84. Franke, A. et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat. Genet. 42, 1118–1125 (2010).
85. Kenny, E. E. et al. A genome-wide scan of ashkenazi jewish crohn's disease suggests novel susceptibility loci. PLoS Genet. 8, (2012).

# Appendices

## Summary

Candidaemia is a bloodstream yeast infection caused by the opportunistic fungal pathogen *Candida albicans* (*C. albicans*). Given the limited arsenal of anti-fungal drugs, together with the emergence of drug resistance, *Candida* is now ranked the fourth most common cause of nosocomial infections[1-4]. However, not all at-risk patients will develop fungal disease, suggesting that inter-individual differences – both genetic and non-genetic factors – play a critical role in determining disease susceptibility. However, it is difficult to identify genetic factors using genome-wide association studies (GWAS) alone because of the limited size of candidaemia patient cohorts and use of inappropriate controls (such as individuals with asymptomatic infections). Given the complexity of human host-*C. albicans* interactions, we followed an integrative approach to obtain a comprehensive picture of these interactions. We integrated multi-omics data (genetics, transcriptomics and proteomics) to identify genes and molecular pathways that are implicated in the anti-*Candida* host immune response. To generate *Candida*-specific data in addition to our candidaemia cohort, we made use of larger population-based cohorts (the 500FG and Lifelines cohorts).

In *chapter 2*, we integrated genetic data from our candidaemia cohort (genotyped by the Immunochip platform) with gene-expression profiles generated upon *Candida* stimulation in peripheral blood mononuclear cells (PBMCs), *Candida*-induced cytokine production capacity in PBMCs (TNFα, IL-6 and IFNγ), and circulating concentrations of three pro-inflammatory cytokines (IL-6, IFNγ and IL-8) in patient serum. Of note, prioritized genes from susceptibility loci were enriched for pathways such as complement and hemostasis, suggesting that the complement system has an important role in disease susceptibility. Given that the Immunochip platform covers 186 loci, we performed the first GWAS to date to identify novel *Candida* susceptibility loci (*chapter 3*). In addition to identifying susceptibility loci, we investigated whether susceptibility can be explained by modulating cytokine levels. To this end, we made use of genetic data and measurements of six cytokines (IL-6, TNFα, IL-1β, IL-17, IL-22 and IFNγ) generated from the 500FG cohort, and mapped the genetic variation that influences cytokine levels in response to *C. albicans*. Interestingly, we observed that susceptibility variants that influence cytokines upon *Candida* stimulation in PBMCs are enriched for lipid and lipoprotein metabolism, suggesting a link between lipid metabolism and inflammation in *Candida* infection. Next, we systematically explored the inflammatory proteins released in candidaemia patients versus healthy individuals by profiling 92 inflammatory proteins in plasma from patients and healthy individuals from the Lifelines cohort (*chapter 4*). In addition to non-genetic factors, we studied the effect of host genetics on the regulation of inflammatory responses by mapping genetic variation that regulates *Candida*-specific inflammatory pro-

teins (protein-QTLs) in a joint analysis of the 500FG and Lifelines DEEP cohorts. Furthermore, we investigated whether pQTLs determine susceptibility to candidaemia and patient survival. We showed a poor enrichment of pQTLs for variants associated to either susceptibility to candidemia or patient survival, which implies a distinct genetic contribution between inflammatory responses, disease susceptibility and patient survival. Lastly, we reviewed advances in laboratory-based and computational approaches for mapping genetic variation in the major histocompatibility complex (MHC) locus, which has been a known risk factor for complex diseases, and discussed the relationship between the MHC variants involved in autoimmune and infectious diseases (*chapter 5*). By comparing SNPs and alleles in the MHC locus associated with both groups of diseases, we observed alleles with opposite allelic effect, supporting the 'pathogen-driven selection' hypothesis, and alleles with common genetic signatures, indirectly supporting the second hypothesis that pathogens can trigger autoimmunity.

In summary, we followed a systems genomics approach to investigate genetics and molecular pathways that contribute to susceptibility to candidaemia. Although there is still a long way to go to develop new drugs, including antifungals, we believe that the currently available technologies, together with advanced computational methods, promise exciting times in the field of systems genomics approaches, not only to candidaemia, but also to other infectious diseases. Although there are still challenges to face before fully embracing genomics in a clinical setting, and challenges to fully unravelling the true risks of disease remain, stratification of patients based on their genetic profile is a promising route toward designing host-directed therapies that would be more effective than the currently available traditional therapeutic approaches.

**References:**

1. Prasad, R., Shah, A. H. & Rawal, M. K. Antifungals: Mechanism of action and drug resistance. Adv. Exp. Med. Biol. 892, 327–349 (2016).

2. Wisplinghoff, H. et al. Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. Clin. Infect. Dis. 39, 309–17 (2004).

3. Wisplinghoff, H. et al. Nosocomial bloodstream infections in pediatric patients in United States hospitals: epidemiology, clinical features and susceptibilities. Pediatr. Infect. Dis. J. 22, 686–691 (2003).

4. Wisplinghoff, H., Seifert, H., Wenzel, R. P. & Edmond, M. B. Current trends in the epidemiology of nosocomial bloodstream infections in patients with hematological malignancies and solid neoplasms in hospitals in the United States. Clin. Infect. Dis. 36, 1103–1110 (2003).

## Samenvatting

Candidaemia is een schimmel infectie in de bloedbaan die veroorzaakt wordt door de opportunistische schimmel pathogeen *Candida albicans* (*C. albicans*). *Candida* staat nu als vierde op de ranglijst voor de meest voorkomende ziekenhuisinfecties, mede door het beperkte repertoire aan anti-schimmel medicatie en nieuw ontwikkelde resistenties tegen medicatie door de pathogeen. Niet iedereen met risico op *Candida* infectie ontwikkelt deze ook daadwerkelijk. Dit suggereert dat verschillen tussen individuen, genetisch of niet-genetisch, een kritische rol spelen in hoe ontvankelijk een persoon is voor candidemia. Het zoeken naar deze factoren wordt bemoeilijkt omdat het lastig is om de genetische factoren voor candidemia ontvankelijkheid te bepalen door een genoomwijde associatiestudie (GWAS), vanwege de beperkte hoeveelheid beschikbare patiënten en het vinden van een gepaste controle group (zoals individuen met asymptomatische infecties). De complexiteit van de interacties tussen mens en *C. albicans* leidde ons ertoe om een integratieve aanpak te kiezen in het vinden van een duidelijk beeld naar al deze interacties. We hebben multi-omics data (genetica, transcriptoom en proteoom) geintegreerd om genen en moleculaire pathways te vinden die geimpliceerd zijn in de immuun respons tegen *Candida*. Om data te genereren die specifiek voor *Candida* is naast ons candideamia cohort, hebben we ook gebruik gemaakt van grotere populatie gebaseerde cohorten (500FG en Lifelines).

In *hoofdstuk 2* integreerden we genetische data van ons candideamia cohort (gegenotypeerd met het Immunochip platform) met genexpressie profielen gegenereerd na *Candida* stimulatie in 'peripheral blood mononuclear cells' (PBMCs), alsook cytokine productie capaciteit in PBMCs na *Candida* inductie (TNFα, IL-6 enIFNγ) en circulerende concentraties van drie pro-inflammatoire cytokinen (IL-6, IFNγ en IL-8) in het serum van patiënten. De genen die geprioriteerd waren uit de loci waren verrijkt in complement en hemostase pathways. Dit suggereert dat het complement systeem een belangrijke rol speelt in candideamia ontvankelijkheid. Het Immunochip platform bestrijkt 186 loci over het genoom, wat ons toeliet om de eerste GWAS te doen en hiermee nieuwe genetische loci voor candideamia susceptibiliteit in kaart te brengen (*hoofdstuk 3*). Naast het identificeren van deze loci, hebben we gekeken of ontvankelijkheid verklaard kan worden door verschillen in cytokine concentraties. Hiervoor hebben we genetische data en metingen van 6 cytokinen (IL-6, TNFα, IL-1β, IL-17, IL-22 en IFNγ) uit het 500FG cohort gecombineerd, om de genetica van cytokine concentraties die te maken hebben met *C. albicans* in kaart te brengen. De varianten die cytokine concentraties veranderen na *Candida* stimulatie in PBMC zijn verrijkt voor lipide en lipoproteïne metabolisme. Dit toont een potentiële link tussen lipide metabolisme en ontstekingsreacties in *Candida* infecties. Hierna hebben we systematisch verkend welke ontstekings eiwitten vrijgegeven worden in candideamia patiënten,

ten opzichte van gezonde individuen. Hiervoor hebben we 92 ontstekings eiwit concentraties geprofileerd uit het plasma van patiënten en gezonde individuen in het lifelines cohort (*hoofdstuk 4*). Naast deze niet-genetische factoren hebben we de genetische factoren van de menselijke gastheer bestudeerd in zijn regulatie van ontstekingsreacties, door de genetische loci in kaart te brengen die *Candida* specifieke ontstekingseiwitten reguleren (pQTL), gecombineerd in twee cohorten: 500FG en Lifelines DEEP. In een verdiepende stap hebben we onderzocht of deze pQTL de ontvankelijkheid voor candidaemia verklaren. pQTL zijn niet verrijkt voor ontvankelijkheid voor candideamia of overlevingskansen van een patiënt. Dit toont dat er waarschijnlijk genetische verschillen bestaan tussen inflammatoire respons en ontvankelijkheid voor en overlevingskansen tijdens candidaemia. Als laatste geven wij een review over de voortgang in laboratorium-gebaseerde en computationele methoden in het in kaart brengen van de genetische regulatie van het `major histocompability complex` (MHC) locus (*hoofdstuk 5*). Het MHC locus is namelijk een bekende risico factor voor complexe ziekten. Hier bespreken we de relatie tussen MHC varianten in autoimmuun- en infectieziekten. Door SNPs en allelen in het MHC locus te vergelijken die met beide typen ziekten geassocieerd zijn. Wij vinden dat allelen tussen deze typen ziekten vaak tegengestelde effecten hebben, wat in lijn is met de `pathogen-driven selection` hypothese. Wij vinden ook allelen met veel voorkomende genetische handtekeningen. Wat ook in lijn is met de tweede hypothese dat pathogenen autoimmuniteit kunnen veroorzaken.

Samenvattend hebben wij systeem genetica technieken toegepast om genetica en moleculaire pathways te onderzoeken die te maken hebben met de ontvankelijkheid van candidaemia. Hoewel er nog een lange weg te gaan is voordat we nieuwe medicatie te kunnen ontwikkelen, waaronder nieuwe anti-schimmel infectie medicatie, geloven wij dat de technieken die nu beschikbaar zijn, gecombineerd met geavanceerde computationele methoden, de belofte hebben van veel succes in het systeem genetica veld. Niet alleen met betrekking tot candideamia, maar ook tot andere infectie ziekten. Er zijn nog veel uitdagingen totdat genomics in de klinische setting toegepast kan worden, alsook het volledig begrijpen van ziekte risico, Maar nu al is het mogelijk om patiënten op te delen aan de hand van hun genetische profiel, om zo gastheer-specifieke behandeling te creeren die effectiever is dan de nu gebruikte traditionele therapie.

## Acknowledgements

I started my Phd research in September 2014 and submitted my PhD book at the end of June 2018. The past years have flown by so fast, and I now find myself writing the last part of this book. Here I want to acknowledge all the people without whom none of my projects would have been possible and without whom I would have not grown to the scientist and person that I am today. It is hard to express my acknowledgement as it is sometimes beyond words.

First and foremost, I would like to thank the members of my thesis assessment committee, Prof. dr. Ingrid Molema, Prof. dr. Hanneke Oude Elberink and Prof. dr. Reinout van Crevel, for carefully reading and evaluating my thesis.

**Department of Genetics**

I would like to thank my promotor, supervisor and mentor, Cisca Wijmenga, for her support and guidance, and for letting me take on challenges and opportunities to complete this research. Every question you asked me during all our meetings opened up new horizons for me to take the research further, to learn (..and learn), and motivated me day-by-day up to this moment. I am so thankful for this. I am also honored that I had the opportunity to attend special events of your life, the awarding of your Spinoza prize, your inauguration to the Lodewijk Sandkuijl Endowed Chair at University Medical Center Groningen, and your highly distinguished appointment of the Knight of the Order of the Netherlands Lion, just to name a few. I'm also thankful for happy moments we shared, such as visits to your house and barbeques with the group. Once again, thank you for giving me the opportunity to join your group and all your direction throughout these years. I wish you all the best in the years to come!

Dear Vinod, you have been a great co-promotor, supervisor and mentor. I cannot thank you enough for giving me the opportunity to work on the genetics of infectious diseases and to turn to bioinformatics during my PhD research. You trusted me and believed that I could manage all challenges. You let me learn and grow as a scientist (and as a person) under your daily supervision and guidance. I wish you all the best in your life!

I would also like to thank Iris. Thank you for supporting me from the beginning of my journey in genetics, for all your feedback, for encouraging me to be more confident, and for taking me to the lab and showing me all your magic on the lab bench. I know that I quit the lab along the way, but you are a great, supportive teacher and person. Always smiling! I wish you the best in your life. Dear Sebo, my research had nothing to do with celiac disease or lincRNAs, but thank you for all feedback, talks and guidance during our meetings. Also, thank you for all fun

group gatherings in your house. All the best in your life!

I would also like to thank the friends and colleagues whose contribution and feedback was invaluable for completing this thesis. Dear Yang, thank you for all the analyses, feedback and advice on how to move on through our challenges, and all the talks throughout these years. Dear Serena, thank you very much for your feedback and tips when needed. I would also like to thank all members of our PhD 'celiac' group: Aarón, Adriaan, Chan, Esteban, Ineke, Joram, Kieu, Maria, Raúl, Renée, Roeland, Olivier and Xiaojing. You are all very nice people and I am very happy and lucky that I met you. Thank you for your input on my projects during our meetings and our talks in the Department and outside of it. You are all great young scientists and I wish you the best in your next steps!

Of course, our work would not have completed without the contribution of the people in our laboratory. Dear Mathieu, Jody, Astrid and Rutger, thank you all so much for your help and little talks throughout these years, Outside the lab, the support from our lovely editors, Jackie and Kate, was also invaluable, Thank you very much both for making our research appropriately readable to the scientific world, including this thesis book! Thank you for all your tips on our writing and on presenting our data, knowledge that I will take with me in the years to come. I cannot thank you enough.

Dear Sasha, thank you for writing, with Cisca and Vinod, the review of this thesis. You were one of the first people that I met in this department as a master's student and one of the inspiring people who motivated me to join the department. I would also like to thank Jingyuan Fu for the data she provided and Lude Franke, Monique and Dylan for our meetings on single-cell sequencing. This list could keep growing forever (Annique, Urmo, Niek, Patrick, Pieter, Alex, Isis…) but, sincerely, I thank all the members of the Department of Genetics for the feedback during my presentations, our daily talks and their contribution to completing this thesis. It was great being member of this group and we can now celebrate the completion of this thesis.

Last but not least, I would like to thank Prof. Jan-Maarten van Dijl from the Department of Medical Microbiology, who invited me to do an internship in his group in August 2011, and encouraged me to apply for the TopMaster in Medical and Pharmaceutical Drug Innovation of the University of Groningen. If I had not applied, I would not probably be in that position today. Thank you for the support.

**Collaborators**

It goes without saying that research is a collaborative effort, and the contribution of each and every one is so valuable for taking each step and gaining a better un-

derstanding on your research questions. I would like to thank our collaborators from the Radboud University whose contribution was significant in doing our research. Mihai Netea, Leo A.B. Joosten, Martin Jaeger, Mark S. Gresnigt, Rob Ter Horst, Marije Oosting, Sanne P. Smeekens, Frank L. van de Veerdonk, Bart Jan Kullberg, thank you all for your contribution! Martin, thank you for all feedback and all the functional experiments that you performed throughout these years. In addition, a thank you to our collaborators from the Duke University Medical Center in North Carolina (United States of America), Melissa D. Johnson and John R. Perfect. Importantly, I would also like to thank all the patients and healthy volunteers who kindly donated blood samples so we were able to perform all the studies described in this thesis.

**Family and Friends**

Completing a thesis takes a lot of time and hard work and, of course, friends to keep you busy and happy when you are not working. I have made great friends during these years in Groningen, both inside and outside of our Department, which makes it difficult for me to fit you all in these lines. Some of you may not be in Groningen any more, but I am thankful that I met you all and you are my friends. Isis, Mamen, Luz and Mireia, you are special ladies, and it was great spending time together. Mireia and Luz, thank for hosting me in Sabadel and Madrid so I had the chance to explore the wonderful cities of Barcelona, Madrid and surroundings. Maria and Andres, thank you for all our talks and Maria for listening to me during our breaks and walks in the city. Dear Kieu, thank you for all the moments we spent together inside and outside the office. Raul, thank you so much for helping me out with any bioinformatic challenge that I came up, at any time, and for our talks. I wish you all the best with Lilly. Lilly, you are very sweet person and it was great meeting you. Arnau, you also taught me some bioinformatics tricks and I cannot thank you enough for your patience. Thank you for being there when my bike was broken, when I had to move, and for everything. Dear Werna, you are a great and fun person, and I am happy that I met you! Thank you for all our talks, dinners and helped me out with moving (together with Arnau). David, thank you for being part of my life, and the trip to Africa. Dear Nshunge, you are like the sister that I did not have. No matter where we are, you will be always in my heart! I know that I have gained a second family by meeting you. Eleni and Mark, I wish you all the best in your life and with your lovely kids, Elisa and Rik. Last but not least, I would like to thank my dear friends back in Greece. We have been together for so many years and have shared so many moments and, even though I am not in Greece any more, you always make me feel the same way I did back in the day. Aunt Roula, thank you for all the walks, talks and laughs especially, every time I came back to Greece for a while.

The last paragraph is dedicated to my family, my mother, father and brother. I

know that you have been missing me every single day, but I know that you are happy when am happy. I feel it is not enough to thank you in a few lines but, any step I made in my life, including the completion of this thesis, is also thanks to you! Aggele, thank you for helping me out to turn this work into a book and for designing my cover! You have done a great job!

Thank you all from the bottom of my heart (including to those whose names were missed in these lines)!

## • Curriculum Vitae

Vasiliki (Vicky) Matzaraki was born in Xanthi, Greece. In 2002 she graduated from high school, passed the High School Exit Examination and was accepted to the Department of Molecular Biology and Genetics of the Democritus University of Thrace in Alexandroupolis, Greece. She graduated with a Bachelor in Molecular Biology and Genetics and continued her studies at the Department of Chemistry of the National and Kapodistrian University of Athens in Greece, where she obtained a Master's degree in Clinical Chemistry (2007-2009). During the second year of her master studies, she joined the Department of Hygiene, Epidemiology and Medical Statistics of the University of Athens Medical School, where she performed epidemiological studies on Human Immunodeficiency Virus (HIV)-1. In 2011, she was awarded an Abel Tasman Talent grant offered by the Graduate School of Medical Sciences of the University of Groningen (the Netherlands), and joined the Molecular Bacteriology Laboratory of the Department of Medical Microbiology of the University Medical Center Groningen (UMCG) for a short internship. During this internship, she studied the expression of surface-exposed proteins of the bacterium *Staphylococcus aureus*.

In 2012, she was accepted to the TopMaster programme Medical and Pharmaceutical Drug Innovation of the Faculty of Medical Sciences of the University of Groningen. During her master studies, Vicky did an internship in the Molecular Bacteriology Laboratory of the Department of Medical Microbiology of the UMCG, where she continued studying the molecular mechanism of staphylococcal spreading motility (January–June 2013). In addition to this, she later joined the Single Molecule Biophysics Laboratory of the Zernike Institute for Advanced Materials of the University of Groningen, where she focused on the kinetic characterization of viral fusion using influenza virus as a model (November 2013–May 2014). She graduated with distinction (*cum laude*) from the TopMaster programme in July 2014 and was awarded a grant from the Graduate School of Medical Sciences to support her three-year PhD research.

In September 2014, she started her PhD research under the supervision of Prof. Cisca Wijmenga and Dr. Vinod Kumar in the Department of Genetics of the UMCG, studying the genetics of infectious diseases with a focus on the bloodstream yeast infection known as candidaemia. During her PhD studies, Vicky was awarded a grant by the Stichting De Cock Foundation to perform functional studies. After completing her PhD, she will start her postdoc research in the group of Prof. Mihai Netea at the Radboud University in Nijmegen, the Netherlands.

## • List of publications

**2018**

Koeken, V.A.C.M., Lachmandas, E., Riza, A., **Matzaraki, V.**, Li, Y., Kumar, V., et al. The role of glutamine metabolism for host defense against *Mycobacterium tuberculosis* infection. The Journal of Infectious Diseases, jiy709, https://doi.org/10.1093/infdis/jiy709

Lachmandas, E., Rios-Miguel, A.B., Koeken, V.A.C.M., van der Pasch, E., Kumar, V., **Matzaraki, V.**, et al. Tissue Metabolic Changes Drive Cytokine Responses to *Mycobacterium tuberculosis*. J. Infect. Dis. 218:165–70.

**2017**

**Matzaraki, V.**, Gresnigt, M.S., Jaeger, M., Ricaño-Ponce, I., Johnson, M.D., Oosting, M., et al. An integrative genomics approach identifies novel pathways that influence candidaemia susceptibility. PLoS One. 12:e0180824.

**Matzaraki, V.,** Kumar, V., Wijmenga, C., Zhernakova, A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome biology. 18(1), 76. doi:10.1186/s13059-017-1207-1

**2016**

Li, Y., Oosting, M., Deelen, P., Ricaño-Ponce, I., Smeekens, S., Jaeger, M., **Matzaraki, V.**, et al. Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. Nat. Med. 22:952–60.

**2014**

Cénit, M.C., **Matzaraki, V.**, Tigchelaar, E.F., Zhernakova, A. Rapidly expanding knowledge on the role of the gut microbiome in health and disease. Biochim. Biophys. Acta - Mol. Basis Dis. 1842:1981–92.